Mathematics for Economists

Ludwig von Auer (Universität Trier)

October 2019

principal textbook: Sydsæter, Hammond, Strøm, Carvajal (2016), Essential Mathematics for Economic Analysis, 5th ed. (older editions are equally suitable)
The book covers our Chapters 1 to 8.

supplementary textbook: Sydsæter, Hammond, Seierstad and Strøm (2008), Further Mathematics for Economic Analysis 2nd. ed. (older edition is equally suitable)
The book covers our Chapter 9.

a very good alternative: Chiang and Wainwright (2005), Fundamental Methods of Mathematical Economics, 4th ed. (older editions are equally suitable)

1 Introductory Topics I: Algebra and Equations

1.1 Some Basic Concepts and Rules

- *natural numbers*:

$$1,\ 2,\ 3,\ 4,\ \ldots$$

- *integers*

$$0,\ \pm 1,\ \pm 2,\ \pm 3,\ \pm 4,\ \ldots$$

  where ± 1 stands for both, $+1$ and -1

- A *real number* can be expressed in the form

$$\pm m.\alpha_1\alpha_2\ldots$$

  Examples of real numbers are

$$-2.5$$
$$273.37827866\ldots$$

Rule

The fraction

$$p/0$$

is not defined for any real number p.

Rule

$$a^{-n} = \frac{1}{a^n}$$

whenever n is a natural number and $a \neq 0$.

- *Warning:*

$$(a+b)^r \neq a^r + b^r$$

Rules of Algebra

(a)	$a+b=b+a$	(g)	$1\cdot a=a$
(b)	$(a+b)+c=a+(b+c)$	(h)	$aa^{-1}=1$, for $a\neq 0$
(c)	$a+0=a$	(i)	$(-a)b=a(-b)=-ab$
(d)	$a+(-a)=0$	(j)	$(-a)(-b)=ab$
(e)	$ab=ba$	(k)	$a(b+c)=ab+ac$
(f)	$(ab)c=a(bc)$	(l)	$(a+b)c=ac+bc$

Rules of Algebra

$$\begin{aligned}(a+b)^2 &= a^2+2ab+b^2 \\ (a-b)^2 &= a^2-2ab+b^2 \\ (a+b)(a-b) &= a^2-b^2\end{aligned} \tag{1}$$

Rules for Fractions

$$
\begin{aligned}
\frac{a \cdot c}{b \cdot c} &= \frac{a}{b} \quad (b \neq 0 \text{ and } c \neq 0) \\
\frac{-a}{-b} &= \frac{(-a)\cdot(-1)}{(-b)\cdot(-1)} = \frac{a}{b} \\
-\frac{a}{b} &= (-1)\frac{a}{b} = \frac{(-1)a}{b} = \frac{-a}{b} \\
\frac{a}{c} + \frac{b}{c} &= \frac{a+b}{c} \\
\frac{a}{b} + \frac{c}{d} &= \frac{a \cdot d}{b \cdot d} + \frac{b \cdot c}{b \cdot d} = \frac{a \cdot d + b \cdot c}{b \cdot d} \\
a + \frac{b}{c} &= \frac{a \cdot c}{c} + \frac{b}{c} = \frac{a \cdot c + b}{c}
\end{aligned}
$$

Rules for Fractions

$$
\begin{aligned}
a \cdot \frac{b}{c} &= \frac{a \cdot b}{c} \\
\frac{a}{b} \cdot \frac{c}{d} &= \frac{a \cdot c}{b \cdot d} \\
\frac{a}{b} \div \frac{c}{d} &= \frac{a}{b} \cdot \frac{d}{c} = \frac{a \cdot d}{b \cdot c}
\end{aligned}
$$

Rules for Powers

$$
\begin{aligned}
a^b a^c &= a^{b+c} \\
\frac{a^b}{a^c} &= a^{b-c} \\
(a^b)^c &= a^{bc} = (a^c)^b \\
a^0 &= 1 \qquad \text{(valid for } a \neq 0\text{, because } 0^0 \text{ is not defined)}
\end{aligned}
$$

- *Remark:* The symbol $\Leftrightarrow$ means "if and only if".

Rule

$$b = c \qquad \Longleftrightarrow \qquad a^b = a^c \tag{2}$$

Rules for Roots

$$
\begin{aligned}
a^{1/2} &= \sqrt{a} \quad (\text{valid if } a \geq 0) \\
\sqrt{ab} &= \sqrt{a}\sqrt{b} \\
\sqrt{\frac{a}{b}} &= \frac{\sqrt{a}}{\sqrt{b}}
\end{aligned}
$$

- Warning:

$$\sqrt{a+b} \neq \sqrt{a} + \sqrt{b}$$

Rules for Roots

$$
\begin{aligned}
a^{1/q} &= \sqrt[q]{a} \\
a^{p/q} &= \left(a^{1/q}\right)^p = \left(a^p\right)^{1/q} = \left(\sqrt[q]{a^p}\right)
\end{aligned}
$$

(p an integer, q a natural number)

Rules for Inequalities

$$
\begin{array}{llll}
a > b & \text{and} \quad b > c & \Rightarrow & a > c \\
a > b & \text{and} \quad c > 0 & \Rightarrow & ac > bc \\
a > b & \text{and} \quad c < 0 & \Rightarrow & ac < bc \\
a > b & \text{and} \quad c > d & \Rightarrow & a + c > b + d
\end{array}
$$

Definition

The *absolute value* of x is denoted by $|x|$, and

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

- Furthermore,

$$|x| \leq a \text{ means that } -a \leq x \leq a$$

1.2 How to Solve Simple Equations

- In the equation
$$3x + 10 = x + 4$$
x is called a *variable*.
- An example with the three variables Y, C and I:
$$Y = C + I$$
- *Solving* an equation means finding all values of the variable(s) that satisfy the equation.

- Two equations that have exactly the same solution are *equivalent equations*.

Rule

To get equivalent equations, do the following to both sides of the equality sign:

- add (or subtract) the same number,
- multiply (or divide) by the same number (different from 0!).

Example

$$
\begin{aligned}
6p - \frac{1}{2}(2p-3) &= 3(1-p) - \frac{7}{6}(p+2) \\
6p - p + \frac{3}{2} &= 3 - 3p - \frac{7}{6}p - \frac{14}{6} \\
6p - p + 3p + \frac{7}{6}p &= \frac{3\cdot 6}{6} - \frac{14}{6} - \frac{3\cdot 3}{2\cdot 3} \\
\frac{8\cdot 6+7}{6}p &= \frac{18-14-9}{6} \\
55p &= -5 \\
p &= \frac{-5}{55} = -\frac{1}{11}
\end{aligned}
$$

Example

$$
\begin{aligned}
\frac{x+2}{x-2}-\frac{8}{x^2-2x} &= \frac{2}{x} && \text{(not defined for } x=2,\ x=0\text{)}\\
\frac{x(x+2)}{x(x-2)}-\frac{8}{x(x-2)} &= \frac{2(x-2)}{x(x-2)} && \text{(for } x\neq 2 \text{ and } x\neq 0\text{)}\\
x(x+2)-8 &= 2(x-2)\\
x^2+2x-8 &= 2x-4\\
x^2 &= 4\\
x &= -2
\end{aligned}
$$

This is the only solution, since for $x=2$ the equation is not defined.

Example

For

$$\frac{z}{z-5} + \frac{1}{3} = \frac{-5}{5-z}$$

no solution exists: For $z \neq 5$ one can multiply both sides by $z-5$ to get

$$\begin{aligned} z + \frac{z-5}{3} &= 5 \\ 3z + z - 5 &= 15 \\ 4z &= 20 \\ z &= 5 \end{aligned}$$

But for $z = 5$ the equation is not defined.

1.3 Equations with Parameters

- Equations can be used to describe a relationship between two variables (e.g., x and y).

Examples

$$\begin{aligned} y &= 10x \\ y &= 3x + 4 \\ y &= -\frac{8}{3}x - \frac{7}{2} \end{aligned}$$

- These equations have a common "linear" structure:

$$y = ax + b$$

where y and x are the variables while a and b are real numbers, called *parameters* or *constants*.

1.4 Quadratic Equations

Definition

Quadratic equations (with one unknow variable) have the general form

$$ax^2 + bx + c = 0 \qquad (a \neq 0) \tag{3}$$

where a, b and c are *constants* (that is, parameters) and x is the unknown variable (for short: the *unknown*)

- Division by the parameter a results in the *equivalent equation*:

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0 \tag{4}$$

Example

Solve the equation

$$x^2 + 8x - 9 = 0$$

The solution applies a method called *completing the square*. This method exploits formula (1)

$$\begin{aligned} x^2 + 8x &= 9 \\ x^2 + 2 \cdot 4 \cdot x &= 9 \\ x^2 + 2 \cdot 4 \cdot x + 4^2 &= 9 + 4^2 \\ (x+4)^2 &= 25 \end{aligned}$$

Therefore, the solutions are $x_1 = 1$ and $x_2 = -9$.

- The general case:

$$
\begin{aligned}
x^2 + \frac{b}{a}x + \frac{c}{a} &= 0 \\
x^2 + \frac{b}{a}x &= -\frac{c}{a} \\
x^2 + 2\left(\frac{b/a}{2}\right)x + \left(\frac{b/a}{2}\right)^2 &= \left(\frac{b/a}{2}\right)^2 - \frac{c}{a} \\
\left(x + \frac{b/a}{2}\right)^2 &= \frac{b^2}{4a^2} - \frac{4ac}{4a^2} \\
4a^2\left(x + \frac{b/a}{2}\right)^2 &= b^2 - 4ac
\end{aligned}
$$

- Note that for

$$b^2 - 4ac < 0$$

no solution would exist.

- However, if $b^2 - 4ac > 0$, the solutions are

$$\begin{aligned} 2a\left(x + \frac{b/a}{2}\right) &= \sqrt{b^2 - 4ac} \\ 2a\left(x + \frac{b/a}{2}\right) &= -\sqrt{b^2 - 4ac} \end{aligned}$$

which is equivalent to

$$2ax + b = \pm\sqrt{b^2 - 4ac} \tag{5}$$

- Solving (5) for x gives the equation on the right hand side of the following rule:

Rule (Quadratic Formula: Version 1)

If $b^2 - 4ac \geq 0$ and $a \neq 0$, then

$$ax^2 + bx + c = 0 \qquad \Leftrightarrow \qquad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{6}$$

The right hand part of (6) is called the *quadratic formula*.

- The quadratic formula could be written also in the form

$$\begin{aligned} x &= \frac{-b/a \pm \sqrt{b^2/a^2 - 4c/a}}{2} \\ &= \frac{-b/a}{2} \pm \frac{\sqrt{b^2/a^2 - 4c/a}}{\sqrt{4}} \\ &= \frac{-b/a}{2} \pm \sqrt{\frac{(b/a)^2 - 4c/a}{4}} \\ &= \frac{-b/a}{2} \pm \sqrt{\frac{(b/a)^2}{4} - c/a} \qquad (7) \end{aligned}$$

- Defining

$$p = \frac{b}{a} \qquad \text{and} \qquad q = \frac{c}{a} \tag{8}$$

equation (4) simplifies to

$$x^2 + px + q = 0 \tag{9}$$

and the quadratic formula (7) to the right hand side of the following rule:

Rule (Quadratic Formula: Version 2)

If $p^2/4 - q \geq 0$, then

$$x^2 + px + q = 0 \qquad \Leftrightarrow \qquad x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q} \tag{10}$$

Example

Consider again the quadratic equation

$$x^2 + 8x - 9 = 0$$

that is, $p = 8$ and $q = -9$. Therefore, the quadratic formula (10) becomes

$$\begin{aligned} x_{1,2} &= -\frac{8}{2} \pm \sqrt{\frac{8^2}{4} + 9} \\ &= -4 \pm \sqrt{16 + 9} \\ &= -4 \pm 5 \end{aligned}$$

and the solutions are

$$x_1 = 1 \qquad \text{and} \qquad x_2 = -9$$

- Another useful rule is:

Rule

If x_1 and x_2 are the solutions of $ax^2 + bx + c = 0$, then

$$ax^2 + bx + c = 0 \qquad \Leftrightarrow \qquad a(x - x_1)(x - x_2) = 0$$

Example

The latter rule implies that

$$x^2 + 8x - 9 = 0$$

with its solutions $x_1 = 1$ and $x_2 = -9$ can be written in the form

$$(x - 1)(x + 9) = 0$$

1.5 Linear Equations in Two Unknowns

- *Economic models* are usually a set of interdependent equations (a *system of equations*).
- The equations of the system can be *linear* or *nonlinear*.
- A (non-economic) example with two linear equations:

$$\begin{aligned} 2x + 3y &= 18 \quad (11) \\ 3x - 4y &= -7 \quad (12) \end{aligned}$$

- We need to find the values of x and y that satisfy *both* equations.

Rule (Method 1)

Solve one of the equations for one of the variables in terms of the other; then substitute the result into the other equation.

Example

From (11)

$$\begin{aligned} 3y &= 18 - 2x \\ y &= 6 - \frac{2}{3}x \end{aligned}$$

Example continued

Inserting in (12) gives

$$\begin{aligned} 3x - 4\left(6 - \frac{2}{3}x\right) &= -7 \\ 3x - 24 + \frac{8}{3}x &= -7 \\ \frac{17}{3}x &= 17 \end{aligned}$$

Dividing both sides by 17 gives

$$\begin{aligned} \frac{1}{3}x &= 1 \\ x &= 3 \end{aligned} \tag{13}$$

Example (continued)

Inserting (13) in (11) gives

$$\begin{aligned} 2 \cdot 3 + 3y &= 18 \\ 3y &= 12 \\ y &= 4 \end{aligned}$$

Rule (Method 2)

Eliminate one of the variables by adding or subtracting a multiple of one equation from the other.

Example

Multiply (11) by 4 and (12) by 3. This gives

$$\begin{aligned} 8x + 12y &= 72 \\ 9x - 12y &= -21 \end{aligned}$$

Then add both equations. This gives

$$\begin{aligned} 17x &= 51 \\ x &= 3 \end{aligned}$$

Inserting this result in (11) gives

$$\begin{aligned} 2 \cdot 3 + 3y &= 18 \\ 3y &= 12 \\ y &= 4 \end{aligned}$$

Rule (Method 3)

Solve both equations for the variable that we want to eliminate first; then set the right hand sides of the two resulting equations equal (or, equivalently, divide one equation by the other, that is, divide the two left hand sides by each other and divide the two right hand sides by each other).

Example

For solving the system

$$y = 5 - x \tag{14}$$
$$-x + y = 1 \tag{15}$$

we solve both equations for y:

$$y = 5 - x \tag{16}$$
$$y = 1 + x \tag{17}$$

Since the left hand sides of (16) and (17) are identical, also the right hand sides are identical and we can write:

$$5 - x = 1 + x \tag{18}$$

Example (continued)

We solve (18) for x:

$$\begin{aligned} 4 &= 2x \\ x &= 2 \end{aligned}$$

Inserting this result in any of the equations (14) to (17) yields

$$y = 3$$

Example

A prominent model from macroeconomics is

$$Y = C + \bar{I} \tag{19}$$
$$C = a + bY \tag{20}$$

where

$$\begin{aligned} Y &= \text{Gross Domestic Product (GDP)} \\ C &= \text{Consumption} \\ \bar{I} &= \text{Investment} \end{aligned}$$

Y and C are considered here as *variables*.
a and b are positive *parameters* of the model with $b < 1$.
Also $\bar{I}$ is a parameter.

Example (continued)

Using *method 1* to solve the macroeconomic model (19) and (20), we first eliminate C by substituting $C = a + bY$ in equation (19):

$$\begin{aligned} Y &= a + bY + \bar{I} \\ Y - bY &= a + \bar{I} \\ (1-b)Y &= a + \bar{I} \\ Y &= \frac{a}{1-b} + \frac{1}{1-b}\bar{I} \qquad (21) \end{aligned}$$

This equation directly tells us for all parameter values (a, b, and $\bar{I}$) the resulting gross domestic product Y.

Example (continued)

Inserting (21) in (20) gives

$$
\begin{aligned}
C &= a + b\left(\frac{a}{1-b} + \frac{1}{1-b}\bar{I}\right) \\
&= \frac{a(1-b)}{1-b} + \frac{ba}{1-b} + \frac{b\bar{I}}{1-b} \\
&= \frac{a + b\bar{I}}{1-b}
\end{aligned}
$$

1.6 Nonlinear Equations

- It is possible also to solve nonlinear equations.
- In the following equations, x, y, z, and w are variables and all other letters are parameters.

Example

The solutions of

$$x^3\sqrt{x+2} = 0$$

are $x = 0$ and $x = -2$.

Example (continued)

The only solutions of

$$x\,(x+a) = x\,(2x+b)$$

are $x = 0$ and $x = a - b$, because for $x \neq 0$ the equation simplifies to

$$x + a = 2x + b$$

which gives the second solution.

The solutions of

$$x\,(y+3)\left(z^2+1\right)\sqrt{w-3} = 0$$

are all x-y-z-w-combinations with $x = 0$ or $y = -3$ or $w = 3$.

Example (continued)

The solutions of

$$\lambda y = \lambda z^2$$

are for $\lambda \neq 0$ all y-z-combinations with $y = z^2$ and for $\lambda = 0$ all y-z-combinations.

2 Introductory Topics II: Miscellaneous

2.1 Summation Notation

- Suppose that there are six regions, each region being denoted by a number:

$$i = 1, 2, 3, 4, 5, 6 \qquad \text{or even shorter} \qquad i = 1, 2, \ldots, 6$$

- Let the population in a region be denoted by N_i. Then the total population of the six regions is

$$N_1 + N_2 + N_3 + N_4 + N_5 + N_6 = N_1 + N_2 + \ldots + N_6 = \sum_{i=1}^{6} N_i$$

- More generally, if there are n regions, the total population is

$$\sum_{i=1}^{n} N_i$$

Examples

$$\begin{aligned}
\sum_{i=1}^{5} i^2 &= 1^2 + 2^2 + 3^2 + 4^2 + 5^2 \\
&= 1 + 4 + 9 + 16 + 25 = 55 \\
\sum_{k=3}^{5} (5k - 3) &= (5 \cdot 3 - 3) + (5 \cdot 4 - 3) + (5 \cdot 5 - 3) = 51 \\
\sum_{i=3}^{n} (x_{ij} - \bar{x}_j)^2 &= (x_{3j} - \bar{x}_j)^2 + (x_{4j} - \bar{x}_j)^2 + \ldots + (x_{nj} - \bar{x}_j)^2
\end{aligned}$$

- The summation sign allows for a compact formulation of lengthy expressions.

Examples

The expression

$$a_1(1-a_1)+a_2(1-a_2)+a_3(1-a_3)+a_4(1-a_4)+a_5(1-a_5)$$

can be written in the compact form

$$\sum_{i=1}^{5} a_i(1-a_i)$$

Examples (continued)

The expression

$$(b)^3 + (2b)^4 + (3b)^5 + (4b)^6 + (5b)^7 + (6b)^8$$

can be written in the compact form

$$\sum_{i=1}^{6} (ib)^{2+i}$$

Rule (Additivity Property)

$$\sum_{i=1}^{n} (a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$$

Rule (Homogeneity Property)

$$\sum_{i=1}^{n} ca_i = c \sum_{i=1}^{n} a_i$$

and if $a_i = 1$ for all i then

$$\sum_{i=1}^{n} ca_i = c \sum_{i=1}^{n} a_i = c\,(n \cdot 1) = cn$$

Rules for Sums

$$\sum_{i=1}^{n} i = 1+2+\ldots+n = \frac{1}{2}n\,(n+1)$$

$$\sum_{i=1}^{n} i^2 = 1^2+2^2+\ldots+n^2 = \frac{1}{6}n\,(n+1)\,(2n+1)$$

$$\sum_{i=1}^{n} i^3 = 1^3+2^3+\ldots+n^3 = \left(\frac{1}{2}n\,(n+1)\right)^2 = \left(\sum_{i=1}^{n} i\right)^2$$

Rule for Sums

$$\sum_{i=0}^{n} a^i = \frac{1-a^{n+1}}{1-a}$$

- Suppose that a firm calculates the total revenues from its sales in Z regions (indexed by i) over S months (indexed by j). The revenues are represented by the rectangular array

$$\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1S} \\ a_{21} & a_{22} & \cdots & a_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ a_{Z1} & a_{Z2} & \cdots & a_{ZS} \end{array}$$

- An element a_{ij} of this array represents the revenues in region i during month j.
- For example, element a_{21} represents the revenues in Region 2 during month 1.

- The total revenues over all S months in some specific region i (the elements in row i) can be written by

$$\sum_{j=1}^{S} a_{ij} = a_{i1} + a_{i2} + \ldots + a_{iS}$$

and the total revenues over all Z regions during some specific month j (the elements in column j) can be written by

$$\sum_{i=1}^{Z} a_{ij} = a_{1j} + a_{2j} + \ldots + a_{Zj}$$

- The total revenues over all Z regions *and* all S months can be expressed by a double sum:

$$\sum_{i=1}^{Z}\left(\sum_{j=1}^{S} a_{ij}\right) = (a_{11}+a_{12}+...+a_{1S})+(a_{21}+a_{22}+...+a_{2S}) +...+(a_{Z1}+a_{Z2}+...+a_{ZS})$$

or equivalently

$$\sum_{j=1}^{S}\left(\sum_{i=1}^{Z} a_{ij}\right) = (a_{11}+a_{21}+...+a_{Z1})+(a_{12}+a_{22}+...+a_{Z2}) +...+(a_{1S}+a_{2S}+...+a_{ZS})$$

- It is usual practice to delete the brackets:

$$\sum_{j=1}^{S}\sum_{i=1}^{Z} a_{ij} = \sum_{i=1}^{Z}\sum_{j=1}^{S} a_{ij}$$

- The double sum notation allows us to write lengthy expressions in a compact way.

Rule

$$\sum_{i=1}^{Z} b_i \sum_{j=1}^{S} a_{ij} b_j = \sum_{i=1}^{Z} \sum_{j=1}^{S} a_{ij} b_i b_j = \sum_{j=1}^{S} \sum_{i=1}^{Z} a_{ij} b_i b_j = \sum_{j=1}^{S} b_j \sum_{i=1}^{Z} a_{ij} b_i$$

Rule

Consider some summation sign $\sum_{i=1}^{Z}$. All variables with index i must be to the right of that summation sign.

Example

Consider the expression

$$
\begin{aligned}
& b_1(a_{11}b_1 + a_{12}b_2 + \dots + a_{1S}b_S) \\
+ \quad & b_2(a_{21}b_1 + a_{22}b_2 + \dots + a_{2S}b_S) \\
& \vdots \\
+ \quad & b_S(a_{S1}b_1 + a_{S2}b_2 + \dots + a_{SS}b_S)
\end{aligned}
$$

This sum can be written in the form

$$\sum_{i=1}^{S} b_i(a_{i1}b_1 + a_{i2}b_2 + \dots + a_{iS}b_S)$$

Example (continued)

Writing the brackets in a more compact form gives

$$\sum_{i=1}^{S} b_i \sum_{j=1}^{S} a_{ij} b_j$$

which can be expressed also in the form

$$\sum_{i=1}^{S} \sum_{j=1}^{S} a_{ij} b_i b_j$$

Example (continued)

Writing the expression

$$\sum_{i=1}^{Z} \sum_{j=1}^{S} a_{ij} b_i b_j$$

in the forms

$$\sum_{i=1}^{Z} b_j \sum_{j=1}^{S} a_{ij} b_i \ , \qquad b_i \sum_{i=1}^{Z} \sum_{j=1}^{S} a_{ij} b_j \ , \text{ or } \qquad \sum_{i=1}^{Z} a_{ij} \sum_{j=1}^{S} b_i b_j$$

is not admissable!

2.2 Essentials of Set Theory

- Suppose that a restaurant serves four different dishes: fish, pasta, omelette, and chicken.
- This menu can be considered as a *set* with four *elements* or *members* (here: dishes):

$$M = \{\text{pasta, omelette, chicken, fish}\}$$

- Notice that the order in which the dishes are listed does not matter.
- The sets

$$A = \{1, 2, 3\} \qquad \text{and} \qquad B = \{3, 2, 1\}$$

are considered *equal*, because each element in A is also in B and each element in B is also in A.

- Sets can contain many other types of elements. For example, the set
$$A = \{(1\,,\,3),(2\,,\,3),(1\,,\,4),(2\,,\,4)\}$$
contains four pairs of numbers.
- Sets could contain infinitely many elements.
- The set of "all" real numbers is denoted by $\mathbb{R}$.
- The set containing as elements "all" pairs of real numbers is denoted by $\mathbb{R}^2$.
- The notation
$$x \in A$$
indicates that the element x *is an element of set* A.
- The notation
$$x \notin A$$
indicates that the element x *is not an element of set* A.

Example

For the set

$$A = \{a, b, c\}$$

one gets $d \notin A$ and for the set

$$B = \mathbb{R}^2$$

one gets $(345.46\,,\,27.42) \in B$.

- Let A and B be any two sets.
- Then A is a *subset* of B if it is true that every member of A is also a member of B.
- Short hand notation: $A \subseteq B$.
- If every member of A is also a member of B and at least one element of B is not in A, then A is a *strict* (or *proper*) *subset* of B: $A \subset B$.
- An empty set $\{ \}$ is denoted by $\varnothing$. The empty set is always a subset of any other set.

Example

The sets

$$A = \{1, 2, 3\} \qquad \text{and} \qquad B = \{1, 2, 3, 4, 5\}$$

give $A \subset B$ and therefore, $A \subseteq B$.
The sets

$$C = \{1, 3, 2, 4\} \qquad \text{and} \qquad D = \{4, 2, 3, 1\}$$

imply that $C \subseteq D$, $D \subseteq C$, and therefore, $C = D$.

- There are three important set operations: union, intersection, and minus.

$A \cup B$ In words: "A union B". The elements that belong to at least one of the sets A and B.

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

$A \cap B$ In words: "A intersection B". The elements that belong to both A and B.

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$A \backslash B$ In words: "A minus B". The elements that belong to A, but not to B.

$$A \backslash B = \{x : x \in A \text{ and } x \notin B\}$$

Example

The sets

$$A = \{1, 2, 3\} \qquad \text{and} \qquad B = \{3, 4, 5\}$$

yield

$$\begin{aligned} A \cup B &= \{1, 2, 3, 4, 5\} \\ A \cap B &= \{3\} \\ A \backslash B &= \{1, 2\} \end{aligned}$$

- Note that

$$A \cap B + A \backslash B = A$$

3 Functions of One Variable

3.1 Basic Definitions

- Suppose that a variable x can take any value from an interval of real values.
- This interval is denoted as the *domain* D of the real variable x.

Definition

A *function* of a real variable x with domain D is a rule that assigns a unique real number to each number x in D.

- As x varies over the whole domain, the set of all possible resulting values $f(x)$ is called the *range* of f.
- Distinguish between the *function* (the rule) f and the *value* $f(x)$ which denotes the value of f at x.

- Functions are often denoted by other letters than f (e.g., g, C, F, ϕ).

Example

$$f(x) = x^3$$

- Often one uses the shorter notation y instead of $f(x)$:

$$y = x^3$$

- y is called the *dependent* (or *endogenous*) variable.
- x is called the *independent* (or *exogenous*) variable.

- The definition of a function is incomplete unless its domain is specified.
- *Convention:* If a function is defined using an algebraic formula, the domain consists of all values of the independent variable for which the formula gives a unique value (unless another domain is explicitly mentioned).

Example

The domain D of

$$f(x) = \frac{1}{x+3}$$

consists of all real numbers $x \neq -3$.

Example

Suppose that the total dollar cost of producing x units of a product is given by

$$C(x) = 100x\sqrt{x} + 500 \tag{22}$$

for each nonnegative real number x that is smaller or equal than the capacity limit x_0: $D = [0, x_0]$. Suppose that $16 < x_0$. The cost of producing $x = 16$ units is

$$\begin{aligned} C(16) &= 100 \cdot 16\sqrt{16} + 500 \\ &= 100 \cdot 16 \cdot 4 + 500 \\ &= 6900 \end{aligned}$$

Definition

A function f is called *increasing* if $x_1 < x_2$ implies $f(x_1) \leq f(x_2)$.

A function f is called *strictly increasing* if $x_1 < x_2$ implies $f(x_1) < f(x_2)$.

A function f is called *decreasing* if $x_1 < x_2$ implies $f(x_1) \geq f(x_2)$.

A function f is called *strictly decreasing* if $x_1 < x_2$ implies $f(x_1) > f(x_2)$.

- The function (22) is strictly increasing.
- The function $f(x) = 4 - 2x$ is strictly decreasing.

3.2 Graphs of Functions

- The *Cartesian coordinate system* (the *x-y-plane*) is useful for depicting functions.
- The *x-axis* together with the *y-axis* separates the plane into four quadrants.
- Any point in the *x-y-plane* represents an *ordered pair* of real numbers (x, y).
- Figure 3-1 depicts the ordered pair $Q = (-5, -2)$ and the ordered pair $P = (3, 4)$.

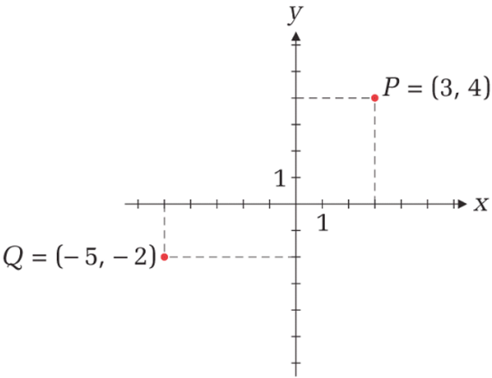


Figure 3-1

- Recall that y is often used as short hand notation for $f(x)$.

Definition

The *graph* of a function f is the set of all points (x, y), where x belongs to the domain of f.

Example

Consider the function

$$y = x^2 - 4x + 3$$

Therefore

x	0	1	2	3	4
y	3	0	-1	0	3

Plotting the points $(0, 3)$, $(1, 0)$, $(2, -1)$, $(3, 0)$, and $(4, 3)$ and then drawing a smooth curve through these points gives the following graph.

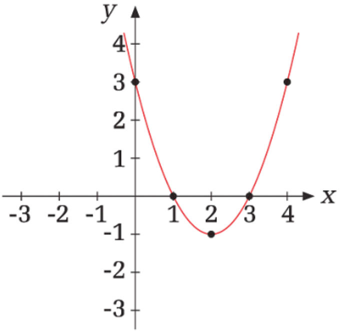


Figure 3-2

- The figure shows a function f with domain D_f and range R_f:

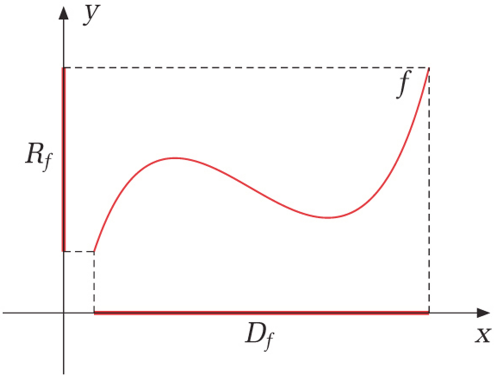


Figure 3-3

- Some important graphs:

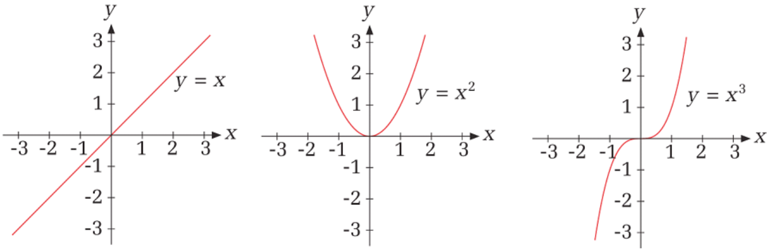


Figure 3-4

- Some other important graphs:

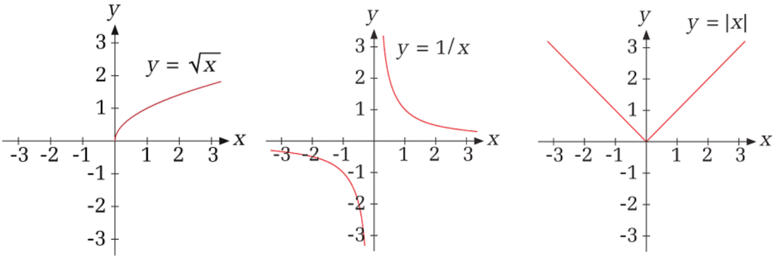


Figure 3-5

3.3 Linear Functions

Definition

A *linear function* has the form

$$f(x) = ax + b$$

with a and b being constants (parameters).

- Take $f(x) = ax + b$ and an arbitrary value of x. Then

$$\begin{aligned} f(x+1) - f(x) &= [a(x+1) + b] - (ax + b) \\ &= ax + a + b - ax - b \\ &= a \end{aligned}$$

- This says that $f(x)$ changes by a units as x is increased by one unit.
- For this reason, the number a is the slope of the graph (a straight line), and so called the *slope* of the linear function.
- If $a > 0$, the line slopes upwards.
- If $a < 0$, the line slopes downwards.
- If $a = 0$, the line is horizontal.
- The absolute value $|a|$ measures the *steepness* of the line.
- Since

$$f(0) = a \cdot 0 + b = b$$

the parameter b indicates the intersection of the graph with the y-axis, that is, the value of $f(x)$ at $x = 0$.

- The lines of linear functions can be used to solve a system of two linear equations in two unknowns.
- This approach corresponds to "Method 3" (see page 36).

Example

A system of two linear equations with two unknowns was given by equations (16) and (17):

$$y = 5 - x \tag{23}$$

$$y = 1 + x \tag{24}$$

Graphically, this system gives the solution point $(x, y) = (2, 3)$; see Figure 3-6.

The algebraic solution gave the same result: $x = 2$ and $y = 3$.

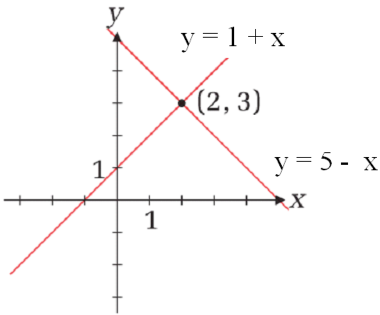


Figure 3-6

3.4 Quadratic Functions

Definition

A *quadratic function* has the form

$$f(x) = ax^2 + bx + c \tag{25}$$

with a, b, and c being constants ($a \neq 0$).

- The graph of such a function is called a *parabola*.
- Its shape roughly resembles $\cup$ when $a > 0$ and $\cap$ when $a < 0$.
- Three typical cases are illustrated in the following diagram (with $b > 0$ and $c > 0$).
- The dashed lines show the *axis of symmetry*.

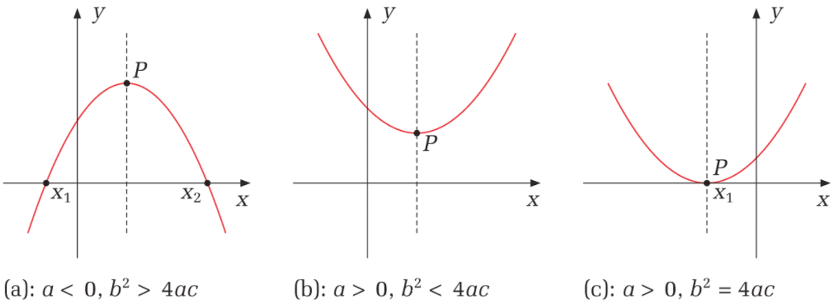


Figure 3-7

- Two key questions:

  1. For which values of x (if any) is
  $$ax^2 + bx + c = 0 \tag{26}$$
  2. What are the coordinates of the maximum/minimum point P (called the *vertex* of the parabola).

- *Answer to Question 1:* If $b^2 - 4ac < 0$, no intersection exists. We know from the quadratic formula (6), that for

$$b^2 - 4ac \geq 0 \tag{27}$$

$$\text{and} \quad a \neq 0 \tag{28}$$

the two x-values

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{29}$$

satisfy (26).

Definition

The values given by the quadratic formula (29) are called the *roots* of the function defined by (25).

- *Answer to Question 2:* The quadratic function yields:

$$\begin{aligned} f(x) &= ax^2 + bx + c \\ &= ax^2 + bx + \frac{b^2}{4a} - \frac{b^2}{4a} + \frac{4ac}{4a} \\ &= a\left(x^2 + 2x\frac{b}{2a} + \frac{b^2}{4a^2}\right) - \frac{b^2}{4a} + \frac{4ac}{4a} \\ &= a\left(x + \frac{b}{2a}\right)^2 - \underbrace{\frac{b^2 - 4ac}{4a}}_{\text{constant}} \end{aligned} \tag{30}$$

- Only the term

$$a\left(x+\frac{b}{2a}\right)^2$$

depends on x.

- The term in brackets is positive except for

$$x=-\frac{b}{2a} \tag{31}$$

- Therefore $f(x)$ reaches a maximum/minimum at (31).
- It is a minimum when $a>0$ and a maximum when $a<0$.

- The axis of symmetry is at position (31).
- From (30) we know that

$$f\left(-\frac{b}{2a}\right) = -\frac{b^2 - 4ac}{4a}$$

Therefore, the vertex P is given by

$$P = \left(-\frac{b}{2a}, -\frac{b^2 - 4ac}{4a}\right)$$

- When $a > 0$ (vertex represents a minimum), then for $b^2 > 4ac$ the vertex is below the x-axis and for $b^2 < 4ac$ the vertex is above the x-axis (then no intersection with the x-axis exists).

Example

The price p per unit obtained by a firm in producing and selling Q units is

$$p(Q) = 102 - 2Q$$

and the cost of producing and selling Q units is

$$C(Q) = 2Q + \frac{1}{2}Q^2$$

Then the profit is

$$\begin{aligned} \pi(Q) &= p(Q) \cdot Q - C(Q) \\ &= (102 - 2Q)\,Q - \left(2Q + \frac{1}{2}Q^2\right) \\ &= -\frac{5}{2}Q^2 + 100Q \qquad (32) \end{aligned}$$

Example continued

Equation (32) is a quadratic function with

$$a = -\frac{5}{2}, \qquad b = 100, \qquad c = 0$$

Since $a < 0$, the profit has a maximum point (rather than a minimum point) at position

$$\begin{aligned} Q &= -\frac{b}{2a} \\ &= -\frac{100}{2(-\frac{5}{2})} \\ &= 20 \end{aligned}$$

Example continued

The corresponding profit is

$$\begin{aligned}\pi(20) &= -\frac{5}{2}20^2 + 100 \cdot 20 \\ &= -1000 + 2000 \\ &= 1000\end{aligned}$$

Using (29), the graph's intersections with the horizontal axis are at

$$Q_1, Q_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-100 \pm \sqrt{100^2}}{-5}$$

which gives $Q_1 = 0$ and $Q_2 = 40$.

3.5 Polynomials

Definition

A *cubic function* has the form

$$f(x) = ax^3 + bx^2 + cx + d \tag{33}$$

with a, b, c, and d being constants ($a \neq 0$).

Example

The graph of

$$f(x) = -x^3 + 4x^2 - x - 6$$

is shown in the following figure.

- Changes in the parameters lead to drastic changes in the graphs.

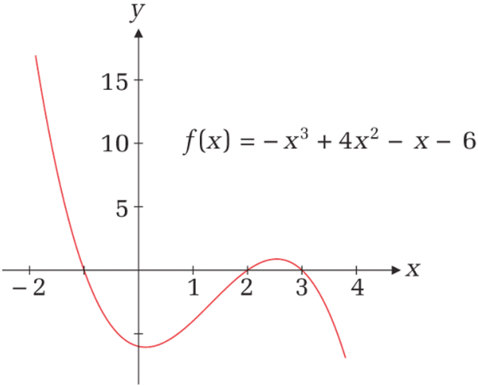


Figure 3-8

- The typical features of a cost function $C(Q)$ are
  - $C(0) > 0$
  - $C(Q)$ strictly increasing in Q
  - starts with a positive but decreasing slope before the slopes starts increasing (as the firm reaches its capacity limit).
- These features require that the parameters in the cost function

$$C(Q) = aQ^3 + bQ^2 + cQ + d$$

  are $a > 0$, $b < 0$, $c > 0$, $d > 0$, and $3ac > b^2$.
- The following graph depicts such a cost function.

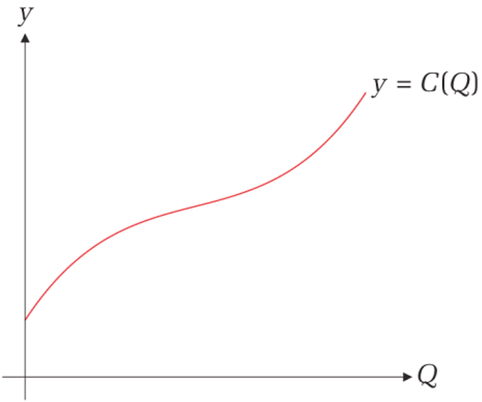


Figure 3-9

Definition

A *general polynomial of degree n* has the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{34}$$

with a_n, a_{n-1}, ..., a_0 being constants ($a_n \neq 0$).

- The equation

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = 0$$

  has *at most n* (real) solutions. That is, the polynomial (34) hast at most n roots.
- Possibly, there are no roots (e.g., $f(x) = x^{100} + 1$).

- The graph corresponding to (34) *has at* most $n-1$ "turning points".

Rule (Fundamental Theorem of Algebra)

Every polynomial of the form (34) can be written as a product of linear and quadratic functions.

3.6 Power Functions

Definition

A *power function* has the form

$$f(x) = Ax^r \tag{35}$$

with $x > 0$, and A and r being constants.

- A special case is $A = 1$:

$$f(x) = x^r \tag{36}$$

- For all r (36) gives $f(1) = 1$.
- The graph corresponding to (36) depends on r (see next figure).

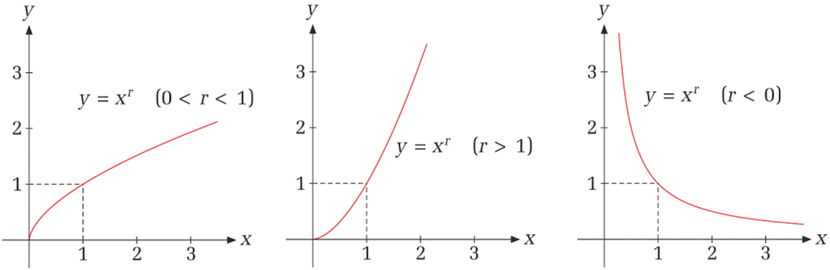


Figure 3-10

3.7 Exponential Functions

- Exponential functions are widely used in statistics and economics.

Definition

An *exponential function* has the form

$$f(x) = Aa^x \tag{37}$$

with A and a being positive constants.

- a is called the *base*.
- Since

$$f(0) = Aa^0 = A$$

(37) can be written in the form

$$f(x) = f(0)a^x$$

- As a consequence

$$f(1) = f(0)a, \qquad f(2) = f(0)a^2 = f(1)a, \qquad \text{etc.}$$

- Therefore, a is the factor by which $f(x)$ increases or decreases when x increases by one unit.
- For $a > 1$ it is an increase and $f(x)$ is stictly increasing.
- For $0 < a < 1$ it is a decrease and $f(x)$ is strictly decreasing.

- A special case is $A = 1$:

$$f(x) = a^x \tag{38}$$

- Note the difference to the power function

$$g(x) = x^a$$

- Since x is often used to describe units of time (periods), it is usually replaced by t:

$$f(t) = Aa^t \tag{39}$$

Rule

A quantity K that increases by $p\%$ per year will have increased after t years to

$$f(t) = K\left(1 + \frac{p}{100}\right)^t$$

A quantity K that decreases by $p\%$ per year will have decreased after t years to

$$f(t) = K\left(1 - \frac{p}{100}\right)^t$$

Example

€ 1000 of savings earning an interest rate of 8% per year ($p = 8$) will have increased after t years to

$$f(t) = 1000 \cdot \left(1 + \frac{8}{100}\right)^t = 1000 \cdot 1.08^t$$

Therefore,

$$\begin{aligned} f(0) &= 1000 \cdot 1.08^0 = 1000 \\ f(1) &= 1000 \cdot 1.08^1 = 1080 \\ &\vdots \\ f(5) &= 1000 \cdot 1.08^5 = 1469.3 \end{aligned}$$

Example

If a car, which at time $t = 0$ has the value A_0, depreciates at the rate of 20% each year, its value $A(t)$ at time t is

$$A(t) = A_0 \left(1 - \frac{20}{100}\right)^t = A_0 0.8^t$$

After 5 years its value is

$$A(5) = A_0 0.8^5 \approx A_0 \cdot 0.32$$

that is, only 32% of its original value.

- In economics and statistics, the most important base a is the (irrational) number $e = 2.718281828459045...$

Definition

The *natural exponential function* has the form

$$f(x) = e^x$$

Rules

All usual rules for powers apply also to this function

$$\begin{aligned} e^s e^t &= e^{s+t} \\ \frac{e^s}{e^t} &= e^{s-t} \\ (e^s)^t &= e^{st} \end{aligned} \tag{40}$$

- Sometimes the notation $\exp(x)$ is used instead of e^x.

3.8 Logarithmic Functions

- If in (39) $a > 1$, how many periods does it take until $f(t)$ doubles (*doubling time*)?
- The value of $f(t)$ in period $t = 0$ is $f(0) = A$.
- We want to know the period t^* such that

$$f(t^*) = 2A$$

that is, we want the value t^* that solves the equation

$$Aa^{t^*} = 2A$$

or more simply, the value of t^* that solves the equation

$$a^{t^*} = 2 \tag{41}$$

- Such questions can be easily answered by using the concept of natural logarithms.
- Let x denote a positive number.

Definition

The *natural logarithm* of x (denoted by $\ln x$) is the power of the number $e(= 2,718...)$ you need to get x:

$$e^{\ln x} = x$$

- More colloquial, $\ln x$ is the answer to the following question:

"e to the power of 'what number' gives x"?

Example

$\ln 1 = 0$, because "e to the power of zero gives 1":

$$e^0 = 1$$

$\ln e = 1$, because "e to the power of 1 gives e":

$$e^1 = e$$

$\ln(1/e) = -1$, because "e to the power of -1 gives $1/e$":

$$e^{-1} = \frac{1}{e}$$

$\ln(e^x) = x$, because "e to the power of x gives e^x":

$$e^x = e^x$$

$\ln(-6)$ is not defined because e^x is positive for all x.

Rules for Natural Logarithms

$$
\begin{aligned}
\ln(xy) &= \ln x + \ln y \\
\ln \frac{x}{y} &= \ln x - \ln y \\
\ln(x^p) &= p \ln x \\
\ln 1 &= 0 \\
\ln e &= 1 \\
e^{\ln x} &= x \\
\ln e^x &= x
\end{aligned} \tag{42}
$$

Rule

$$\text{for } x > 0,\ y > 0: \qquad x = y \qquad \Longleftrightarrow \qquad \ln x = \ln y$$

- Warning:

$$\ln(x+y) \neq \ln x + \ln y$$

- What is the solution to (41)? (41) is equivalent to

$$\begin{aligned} \ln\left(a^{t^*}\right) &= \ln 2 \\ t^* \ln a &= \ln 2 \\ t^* &= \frac{\ln 2}{\ln a} \end{aligned}$$

Definition

The function

$$f(x) = \ln x$$

is called the *natural logarithmic function* of x. Its domain is $x > 0$.

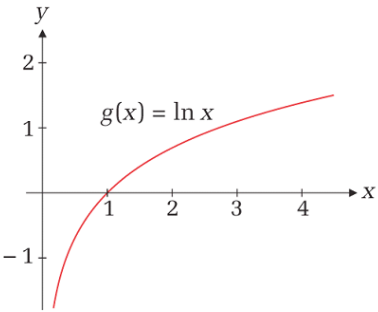


Figure 3-11

- Also logarithms based on numbers other than e exist.

Definition

The *logarithm of x to base a* (denoted by $\log_a x$) is the power of the base a you need to get x:

$$a^{\log_a x} = x$$

- More colloquial, $\log_a x$ is the answer to the following question:

"a to the power of 'what number' gives x"?

Example

$$\log_2 8 = 3$$

Rules

The same rules as for the natural logarithm apply:

$$\begin{aligned}
\log_a(xy) &= \log_a x + \log_a y \\
\log_a \frac{x}{y} &= \log_a x - \log_a y \\
\log_a(x^p) &= p \log_a x \\
\log_a 1 &= 0 \\
\log_a a &= 1
\end{aligned}$$

3.9 Shifting Graphs

- This section studies in general how the graph of a function $f(x)$ relates to the graphs of the functions

$$f(x) + c, \qquad f(x + c), \qquad \text{and} \qquad cf(x),$$

where c is positive constant.

- As an example, the function

$$y = \sqrt{x}$$

is considered.

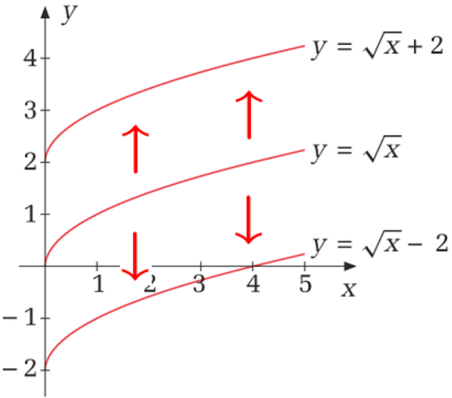


Figure 3-12

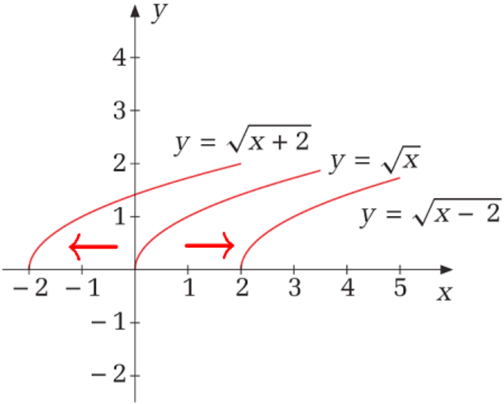


Figure 3-13

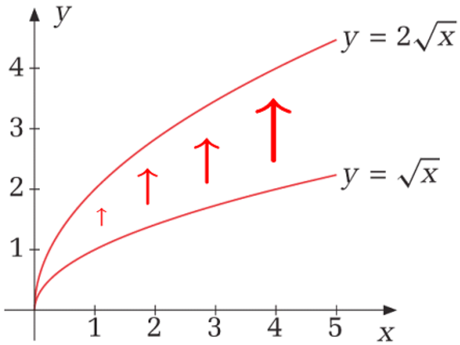


Figure 3-14

Rule

(i) If $y = f(x)$ is replaced by $y = f(x) + c$, the graph is moved upwards by c units if $c > 0$ (downwards if c is negative).

(ii) If $y = f(x)$ is replaced by $y = f(x + c)$, the graph is moved c units to the left if $c > 0$ (to the right if c is negative).

(iii) If $y = f(x)$ is replaced by $y = cf(x)$, the graph is stretched vertically if $c > 1$ and compressed if $0 < c < 1$(stretched or compressed vertically and reflected about the x-axis if c is negative).

- As a result, the graph of the function

$$y = 2 - (x+2)^2$$

can be constructed with the graph of $y = x^2$ as a reference.
- The graph of $y = x^2$ can be
  1. reflected about the x-axis,
  2. moved to the left by two units, and finally
  3. moved upwards by two units.
- Other sequences of these three steps are equally fine.

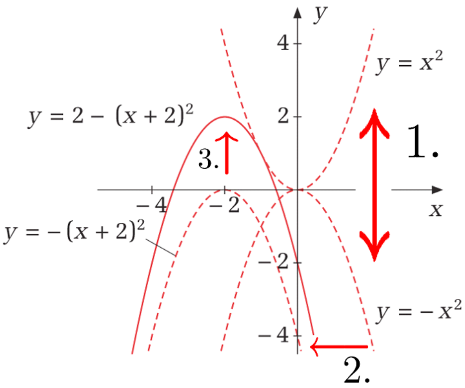


Figure 3-15

3.10 Computing With Functions

- Let $f(t)$ and $m(t)$ denote the number of female and male students in year t, while $n(t)$ denotes the total number of students.
- Then
$$n(t) = f(t) + m(t)$$
- The graph of $n(t)$ is obtained by piling the graph of $f(t)$ on top of the graph of $m(t)$.

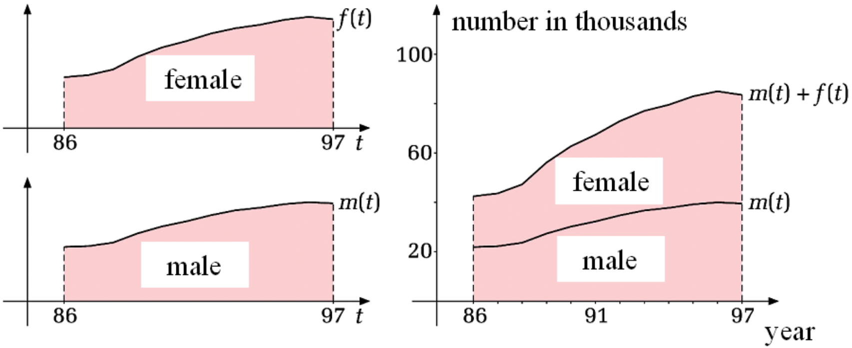

f(t)
female
86
97 t
m(t)
male
86
97 t
number in thousands
100
60
20
m(t) + f(t)
female
m(t)
male
86
91
97
year

Figure 3-16

- Suppose that f and g are functions which both have the same domain, namely an interval in the set of real numbers.
- The sum of f and g is also a function. Here this function is denoted as h

$$h(x) = f(x) + g(x)$$

- The difference between f and g is also a function. Here this function is denoted as k

$$k(x) = f(x) - g(x)$$

Example

When the cost function is

$$C(Q) = aQ^3 + bQ^2 + cQ + d$$

the average cost function is

$$\begin{aligned} A(Q) &= \frac{aQ^3 + bQ^2 + cQ + d}{Q} \\ &= aQ^2 + bQ + c + \frac{d}{Q} \end{aligned}$$

This is the sum of a quadratic function $(aQ^2 + bQ + c)$ and a so-called hyperbolic function (d/Q).

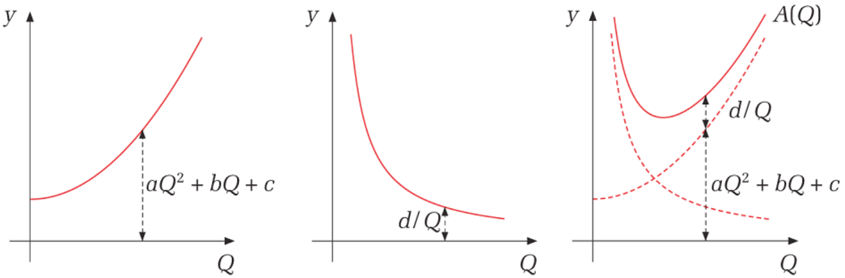


Figure 3-17

Example

Let $R(Q)$ denote the revenues obtained by producing and selling Q units and suppose that the firm gets a fixed price p per unit.

Therefore $R(Q)$ is a straight line through the origin.

The profit $\pi(Q)$ is given by

$$\pi(Q) = R(Q) - C(Q)$$

The graph of $-C(Q)$ must be added to $R(Q)$. This is equivalent to subtracting the graph $C(Q)$ from $R(Q)$.

The maximum profit is at output Q^*.

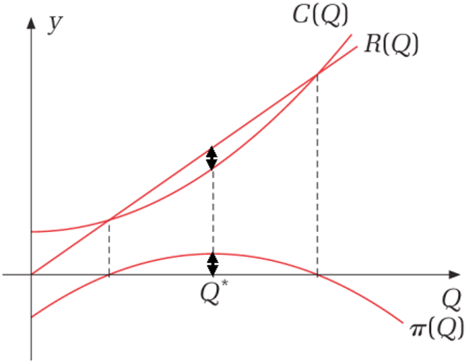


Figure 3-18

- Suppose that f and g are functions which both are defined in a set A of real numbers.
- The product of f and g is also a function. Here this function is denoted as h

$$h(x) = f(x) \cdot g(x)$$

- The quotient of f and g is also a function. Here this function is denoted as k

$$k(x) = \frac{f(x)}{g(x)}$$

with $g(x) \neq 0$.

Definition

Suppose that $y = f(u)$ and $u = g(x)$. Then y is a *composite function* of x:

$$y = f(g(x))$$

with

$g(x)$ being the *interior function* (or *kernel*) and

f being the *exterior function.*

- The composite function $y = f(g(x))$ is often denoted by $f \circ g$ and it is read as "f of g" or "f after g".
- $f \circ g$ and $g \circ f$ are very different composite functions.
- Do not confuse $f \circ g$ with $f \cdot g$.

Example

Consider the composite function

$$y = e^{-(x-\mu)^2}$$

with μ being a constant.

The choice of the interior and exterior function is to some degree arbitrary.

One could define $g(x) = -(x-\mu)^2$ as the interior function and $f(u) = e^u$ as the exterior function.

Alternatively, one could define $g(x) = (x-\mu)^2$ as the interior function and $f(u) = e^{-u}$ as the exterior function.

3.11 Inverse Functions

- Suppose that the demand quantity D for a commodity depends on the price per unit P according to

$$D = \frac{30}{P^{1/3}} \tag{43}$$

- This gives for $P = 27$ the demand quantity

$$D = \frac{30}{27^{1/3}} = \frac{30}{3} = 10$$

- From the perspective of the producers, however, it may be more natural to treat output as something that the producer can choose and to consider the resulting price.

- For this purpose (43) must be *inverted*, that is, P must become a function of D:

$$\begin{aligned} P^{1/3}D &= 30 \\ P^{1/3} &= \frac{30}{D} \\ \left(P^{1/3}\right)^3 &= \left(\frac{30}{D}\right)^3 \\ P &= \frac{27000}{D^3} \qquad (44) \end{aligned}$$

- (44) is the *inverse function* of (43).
- Solving (44) for D, that is, inverting (44) gives (43).
- Therefore, (43) and (44) are inverse functions of each other, or more simply, *inverses*.
- Both functions convey exactly the same information.

- Let f be a function with domain D_f.
- This says that to each x in D_f there corresponds a unique number $f(x)$.
- Then the range of f is R_f and consists of all numbers $f(x)$ obtained by letting x vary in D_f.

Definition

The function f is said to be *one-to-one in* D_f if f never has the same value at any two different points in D_f.

- Then for each one y in R_f there is exactly one x in D_f such that $y = f(x)$.
- The following diagram shows on the left a function f that is one-to-one in D_f and on the right a function g that is not one-to-one in D_f.

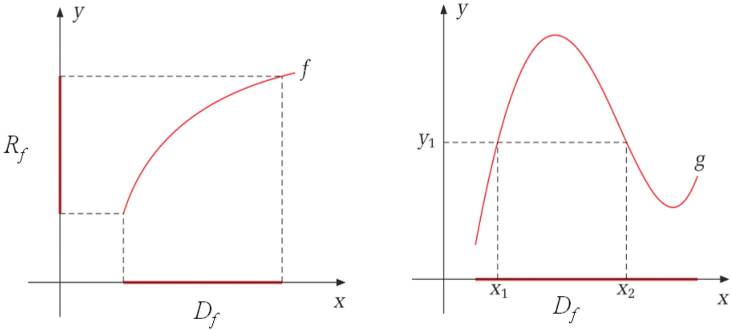


Figure 3-19

- Let f be a function with domain D_f and range R_f.

Rule

If and only if f is one-to-one, it has an inverse function g with domain $D_g = R_f$ and range $R_g = D_f$. This function g is given by the following rule: For each y in D_g the value $g(y)$ is the unique number x in R_g such that $f(x) = y$. Then

$$g(y) = x \qquad \Longleftrightarrow \qquad y = f(x)$$

with x in D_f and y in D_g.

- As a direct implication

$$g(f(x)) = x$$

In words: g undoes what f did to x.

Rule

If g is the inverse function of f, then f is the inverse function of g and vice versa.

- If g is the inverse function of f, it is standard to use the notation f^{-1} for g.
- Note that f^{-1} does not mean $1/f$!

Rule

The inverse of the natural exponential function

$$y = e^x$$

is the natural logarithmic function

$$x = \ln y$$

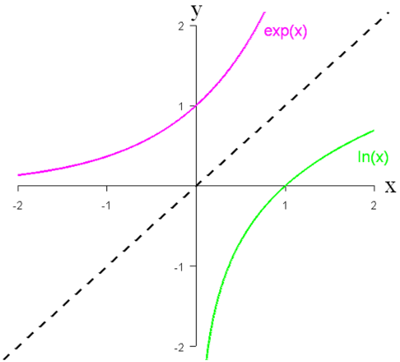


Figure 3-20

4 Differentiation

4.1 Slopes of Curves

- For the graph representing the function $y = ax + b$ the slope was given by the number a.
- Consider some arbitrary function f.
- The slope of the corresponding graph at some point x_0 is the slope of the tangent to the graph at x_0.
- In Figure 4-1, point P has the coordinates $(x_0, f(x_0))$.
- The straight line T is the tangent line to the graph at point P.
- It just touches the curve at point P.
- The slope of the graph at x_0 is the slope of T.
- This slope is $1/2$.

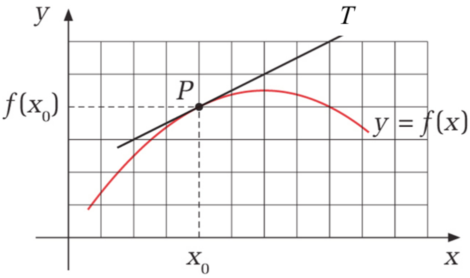


Figure 4-1

4.2 Tangents and Derivatives

Definition

The slope of the tangent line at point $(x, f(x))$ is called the *derivative* of f at point x. This number is denoted by $f'(x)$.

- Read $f'(x)$ as "f prime x".
- In Figure 4-1 the point $x = x_0$ was considered.
- The derivative of f at point x_0 was

$$f'(x_0) = \frac{1}{2}$$

- In Figure 4-2, P and Q are points on the curve (graph).
- The entire straight line through P and Q is called a *secant*.
- Keep P fixed, but move Q along the curve towards P.
- Then the secant rotates around P towards the limiting straight line T.
- T is the *tangent (line)* to the curve at P.

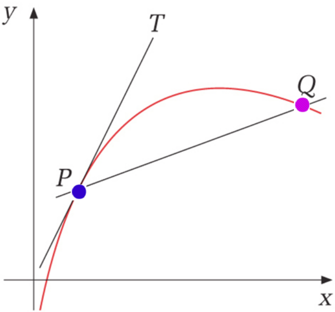


Figure 4-2

- Define Δx to be the distance between x_0 and the x-coordinate of point Q (see Figure 4-3).
- The coordinates of the points P and Q can be written in the form

$$P = (\,x_0\,,\,f(x_0)\,) \qquad \text{and} \qquad Q = (\,x_0 + \Delta x\,,\,f(x_0 + \Delta x)\,)$$

- The slope m_{PQ} of the secant PQ is

$$\begin{aligned} m_{PQ} &= \frac{f(x_0 + \Delta x) - f(x_0)}{(x_0 + \Delta x) - x_0} \\ &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \end{aligned}$$

- For $\Delta x = 0$ this quotient is not defined.

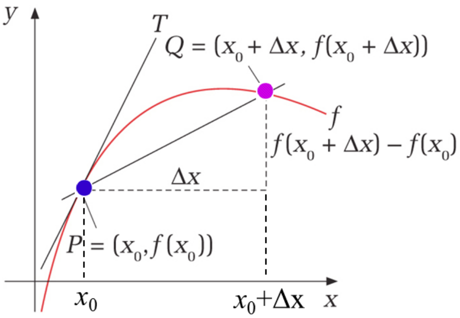


Figure 4-3

- As Q moves towards P, Δx tends to 0 and the slope of the secant PQ tends towards the slope of the tangent T.
- The mathematical symbol

$$\lim_{\Delta x \to 0}$$

in front of some expression denotes the value of the expression as Δx tends towards 0.

Definition

The derivative of the function f at point x_0, denoted by $f'(x_0)$, is given by the formula

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{45}$$

Example

The derivative of $f(x) = x^2$ at point x_0 is according to formula (45)

$$\begin{aligned} f'(x_0) &= \lim_{\Delta x \to 0} \frac{(x_0 + \Delta x)^2 - (x_0)^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{(x_0)^2 + 2x_0 \Delta x + (\Delta x)^2 - (x_0)^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{2x_0 \Delta x + (\Delta x)^2}{\Delta x} \end{aligned}$$

For all $\Delta x \neq 0$ we can cancel Δx and obtain

$$f'(x_0) = \lim_{\Delta x \to 0} (2x_0 + \Delta x) = 2x_0$$

- By $f'(x)$ we mean the function that gives us for every point x_0 the derivate of $f(x)$ at point x_0.
- We call $f'(x)$ the *derivative* of $f(x)$.

- In place of $f'(x)$ often y' or the *differential notation* of Leibniz is used:

$$\frac{\mathrm{d}y}{\mathrm{d}x}, \qquad \mathrm{d}y \,/\, \mathrm{d}x, \qquad \frac{\mathrm{d}f(x)}{\mathrm{d}x}, \qquad \mathrm{d}f(x) \,/\, \mathrm{d}x, \qquad \frac{\mathrm{d}}{\mathrm{d}x}f(x)$$

- The derivative $f'(x)$ can be used to define the notion of increasing and decreasing functions.

Definition

$$\begin{array}{lcl} f'(x) \geq 0 \text{ for all } x \text{ in } D_f & \iff & f \text{ is increasing in } D_f \\ f'(x) > 0 \text{ for all } x \text{ in } D_f & \iff & f \text{ is strictly increasing in } D_f \\ f'(x) \leq 0 \text{ for all } x \text{ in } D_f & \iff & f \text{ is decreasing in } D_f \\ f'(x) < 0 \text{ for all } x \text{ in } D_f & \iff & f \text{ is strictly decreasing in } D_f \end{array}$$

4.3 Rules for Differentiation

- The derivative of a function f at point x_0 was defined by

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Definition

If this limit exists, f is *differentiable* at x_0. If f is differentiable at every point x_0 in the domain D_f, then we call f *differentiable*.

Rules of Differentiation

Rule 1:	$f(x) = A$	$\Rightarrow$	$f'(x) = 0$
Rule 2:	$f(x) = A + g(x)$	$\Rightarrow$	$f'(x) = g'(x)$
Rule 3:	$f(x) = Ag(x)$	$\Rightarrow$	$f'(x) = Ag'(x)$

Examples

$$
\begin{aligned}
f(x) &= 5 & \Rightarrow \quad & f'(x) = 0 \\
f(x) &= 5 + 2x & \Rightarrow \quad & f'(x) = 2 \\
f(x) &= 5 \cdot 2x & \Rightarrow \quad & f'(x) = 5 \cdot 2 = 10
\end{aligned}
$$

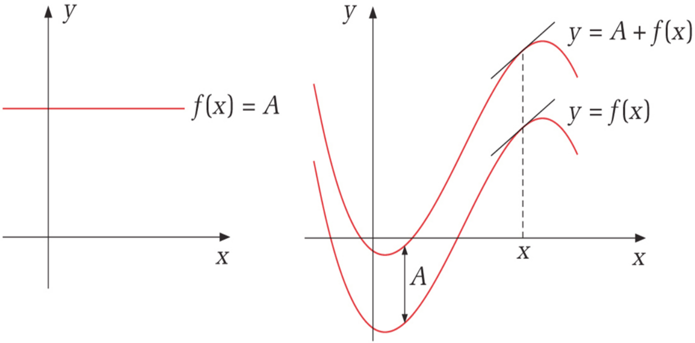


Figure 4-4

Rule of Differentiation

Rule 4 (power rule): $f(x) = x^a \quad \Rightarrow \quad f'(x) = ax^{a-1}$

with a being an arbitrary constant.

Examples

$$\begin{aligned} f(x) &= x^3 \quad \Rightarrow \quad f'(x) = 3x^2 \\ f(x) &= 3x^8 \quad \Rightarrow \quad f'(x) = 3 \cdot 8x^7 = 24x^7 \end{aligned}$$

Rule of Differentiation

Rule 5 (sums): If both f and g are differentiable at x, then the sum $f+g$ and the difference $f-g$ are both differentiable at x, and

$$h(x) = f(x) \pm g(x) \qquad \Rightarrow \qquad h'(x) = f'(x) \pm g'(x)$$

Example

$$h(x) = x^3 - 5x^{-2} \qquad \Rightarrow \qquad \begin{aligned} h'(x) &= 3x^2 - \left(-2 \cdot 5x^{-3}\right) \\ &= 3x^2 + 10x^{-3} \end{aligned}$$

Rule of Differentiation

Rule 6 (products): If both f and g are differentiable at x, then so is $h = f \cdot g$, and

$$h(x) = f(x) \cdot g(x) \qquad \Rightarrow \qquad h'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Example

The function

$$h(x) = \left(x^3 - x\right)\left(5x^4 + x^2\right)$$

can be written as

$$h(x) = f(x) \cdot g(x)$$

with

$$\begin{aligned} f(x) &= \left(x^3 - x\right) \\ g(x) &= \left(5x^4 + x^2\right) \end{aligned}$$

Therefore

$$\begin{aligned} h'(x) &= f'(x) \cdot g(x) + f(x) \cdot g'(x) \\ &= \left(3x^2 - 1\right)\left(5x^4 + x^2\right) + \left(x^3 - x\right)\left(20x^3 + 2x\right) \\ &= 35x^6 - 20x^4 - 3x^2 \end{aligned}$$

Rule of Differentiation

Rule 7 (quotient): If both f and g are differentiable at x and $g(x) \neq 0$, then $h = f/g$ is differentiable at x, and

$$h(x) = \frac{f(x)}{g(x)} \qquad \Rightarrow \qquad h'(x) = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{\left(g(x)\right)^2}$$

Example

The derivative of the function

$$h(x) = \frac{3x-5}{x-2} = \frac{f(x)}{g(x)}$$

is

$$\begin{aligned} h'(x) &= \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{(g(x))^2} \\ &= \frac{3\cdot(x-2)-(3x-5)\cdot 1}{(x-2)^2} \\ &= \frac{-1}{(x-2)^2} \end{aligned}$$

Note that $h(x)$ is strictly decreasing at all $x \neq 2$.

Rule of Differentiation

Rule 8 (chain rule): If g is differentiable at x and f is differentiable at $u = g(x)$, then the composite function $h(x) = f(g(x))$ is differentiable at x, and

$$h'(x) = f'(u) \cdot g'(x) = f'(g(x)) \cdot g'(x)$$

- In words: First differentiate the exterior function with respect to the interior function (kernel), then multiply by the derivative of the interior function.

Example

Let $f(u) = u^3$ and $g(x) = 2 - x^2$. The derivative of

$$h(x) = f(g(x)) = \left(2 - x^2\right)^3$$

is

$$\begin{aligned} h'(x) &= f'(g(x)) \cdot g'(x) \\ &= 3\left(2 - x^2\right)^2 \cdot (-2x) \\ &= -6x\left(4 - 4x^2 + x^4\right) \\ &= -6x^5 + 24x^3 - 24x \end{aligned}$$

- Expressing the eight rules in Leibniz's differential notation gives

$$
\begin{aligned}
\text{Rule 1} &: & \frac{\mathrm{d}}{\mathrm{d}x}A = 0 \\
\text{Rule 2} &: & \frac{\mathrm{d}}{\mathrm{d}x}\left[A + f(x)\right] = \frac{\mathrm{d}}{\mathrm{d}x}f(x) \\
\text{Rule 3} &: & \frac{\mathrm{d}}{\mathrm{d}x}\left[Af(x)\right] = A\frac{\mathrm{d}}{\mathrm{d}x}f(x) \\
\text{Rule 4} &: & \frac{\mathrm{d}}{\mathrm{d}x}\left(x^a\right) = ax^{a-1} \\
\text{Rule 5} &: & \frac{\mathrm{d}}{\mathrm{d}x}\left[f(x) \pm g(x)\right] = \frac{\mathrm{d}}{\mathrm{d}x}f(x) \pm \frac{\mathrm{d}}{\mathrm{d}x}g(x)
\end{aligned}
$$

Rule 6 : $$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)\cdot g(x)\right] = \left[\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right]\cdot g(x) + f(x)\cdot\left[\frac{\mathrm{d}}{\mathrm{d}x}g(x)\right]$$

Rule 7 : $$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{f(x)}{g(x)}\right] = \frac{\left[\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right]\cdot g(x) - f(x)\cdot\left[\frac{\mathrm{d}}{\mathrm{d}x}g(x)\right]}{g(x)^2}$$

Rule 8 : $$\frac{\mathrm{d}}{\mathrm{d}x}f\left(g\left(x\right)\right) = \frac{\mathrm{d}}{\mathrm{d}g\left(x\right)}f\left(g\left(x\right)\right)\cdot\frac{\mathrm{d}}{\mathrm{d}x}g\left(x\right)$$

4.4 Higher-Order Derivatives

- The derivate f' of a function $y = f(x)$ is called the *first derivate* of f.
- If f' is also differentiable, then we can differentiate f' in turn.
- The result is called the *second order derivative* and it is written as f'' or y''.

Definition

$f''(x)$ is the second order derivative of f evaluated at the particular point x.

- f'' or y'' can be written in the differential notation as

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right]$$

or more simply as

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \qquad \text{or} \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}$$

Example

The first derivative of

$$f(x) = 2x^5 - 3x^3 + 2x$$

is

$$f'(x) = 10x^4 - 9x^2 + 2$$

Therefore, the second order derivative is

$$f''(x) = 40x^3 - 18x$$

- Let I denote some interval on the real line.
- The second order derivative $f''(x)$ is the derivative of $f'(x)$. Therefore

$$\begin{aligned} f''(x) &\geq 0 \text{ on } I &\iff&& f' \text{ is increasing on } I \\ f''(x) &\leq 0 \text{ on } I &\iff&& f' \text{ is decreasing on } I \end{aligned}$$

- The consequences are illustrated in the following figure.

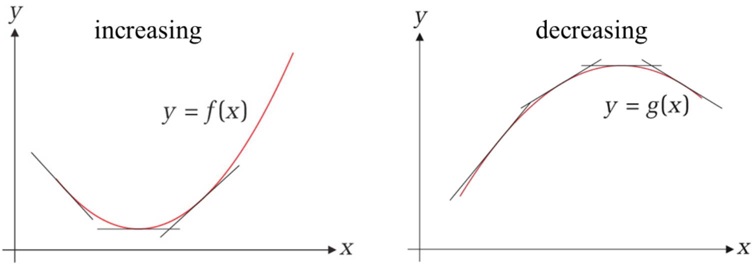


Figure 4-5

- Suppose that f is continuous in the interval I and twice differentiable in the interior of I.

Definition

$$\begin{array}{lclcl} f''(x) & \geq & 0 \text{ for all } x \text{ in } I & \iff & f \text{ is \textit{convex} on } I \\ f''(x) & \leq & 0 \text{ for all } x \text{ in } I & \iff & f \text{ is \textit{concave} on } I \end{array}$$

- If I is the real line, the interval is not mentioned explicitly ("f is convex" or "f is concave").
- One can further distinguish between *increasing convex* and *decreasing convex* and also between *increasing concave* and *decreasing concave* (see next figure).

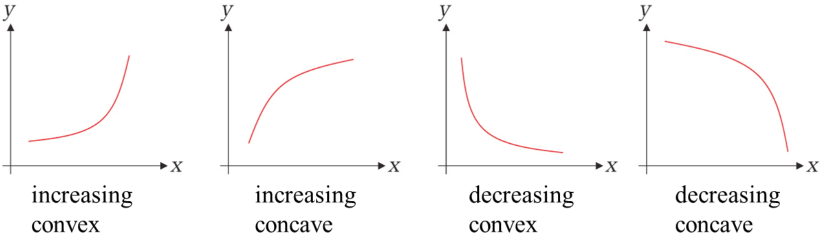

y
x
increasing
convex
y
x
increasing
concave
y
x
decreasing
convex
y
x
decreasing
concave

Figure 4-6

- Let $y = f(x)$. The derivate of f'' is called the third-order derivative and is denoted by

$$f''' \qquad \text{or} \qquad y''' \qquad \text{or} \qquad \frac{\mathrm{d}^3}{\mathrm{d}x^3} f(x)$$

- Correspondingly, the nth derivative of f is denoted by

$$f^{(n)} \qquad \text{or} \qquad y^{(n)} \qquad \text{or} \qquad \frac{\mathrm{d}^n}{\mathrm{d}x^n} f(x)$$

4.5 Derivative of the Exponential Function

- The derivative of a function f was defined by

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

- For the natural exponential function $f(x) = e^x$ this definition gives (note that e^{x_0} is a constant):

$$\begin{aligned} f'(x_0) &= \lim_{\Delta x \to 0} \frac{e^{x_0 + \Delta x} - e^{x_0}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{e^{x_0} e^{\Delta x} - e^{x_0}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{e^{x_0} \left(e^{\Delta x} - 1 \right)}{\Delta x} \\ &= e^{x_0} \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} \end{aligned}$$

- It can be shown that

$$\lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} = 1$$

- Therefore,

$$f'(x_0) = e^{x_0} \cdot 1 = e^{x_0}$$

Rule of Differentiation

Rule 9:

$$f(x) = e^x \qquad \Rightarrow \qquad f'(x) = e^x$$

The derivative of $f(x) = e^x$ is equal to the function itself.

- Since

$$f(x) = e^x > 0$$

the same is true for the derivative $f'(x)$.

- Rule 9 can be combined with the chain rule (rule 8):

$$f(x) = e^{g(x)} \qquad \Rightarrow \qquad f'(x) = e^{g(x)} g'(x)$$

Example

The derivative of

$$f(x) = x^p e^{ax} \qquad \text{(with } p \text{ and } a \text{ being constants)}$$

is (exploiting the product rule and the chain rule)

$$\begin{aligned} f'(x) &= px^{p-1}e^{ax} + x^p e^{ax} a \\ &= px^{p-1}e^{ax} + x^{p-1}x^1 e^{ax} a \\ &= x^{p-1}e^{ax}(p + ax) \end{aligned}$$

- The derivative of

$$f(x) = a^x$$

  with a being some positive constant can be computed by exploiting rule 9.
- Using (40) and (42), we get

$$f(x) = a^x = \left(e^{\ln a}\right)^x = e^{(\ln a)x}$$

  Therefore, the chain rule gives

$$f'(x) = e^{(\ln a)x} \ln a = a^x \ln a \tag{46}$$

- Note that for $a = e$ the derivative simplifies to $f'(x) = e^x$.
- Therefore, (46) is a generalisation of rule 9.

Example

The derivative of

$$f(x) = x2^{3x} = x\left(2^3\right)^x = x8^x$$

is, using the product rule and (46),

$$\begin{aligned} f'(x) &= 8^x + x8^x \ln 8 \\ &= 8^x (1 + x \ln 8) \end{aligned}$$

4.6 Derivative of the Natural Logarithmic Function

- The natural logarithmic function is

$$g(x) = \ln x$$

- Due to (2) it is equivalent to

$$e^{g(x)} = e^{\ln x}$$

  and, using (42), to

$$e^{g(x)} = x \tag{47}$$

- The left and right-hand sides of this equation can be considered as two functions of x, namely $h(x) = e^{g(x)}$ and $k(x) = x$. At all values of x these two functions have the same value (that is, their graphs are identical).

- Therefore, also the derivatives, $h'(x)$ and $k'(x)$, have the same value.
- Differentiating both sides of (47) with respect to x gives

$$e^{g(x)}g'(x) = 1 \tag{48}$$

- Making use of (47), (48) can be written in the form

$$g'(x) = \frac{1}{x}$$

giving rise to the following rule:

Rule of Differentiation

Rule 10: $f(x) = \ln x \quad \Rightarrow \quad f'(x) = \frac{1}{x}$

- Combining rule 10 and the chain rule gives

$$f(x) = \ln g(x) \qquad \Rightarrow \qquad f'(x) = \frac{1}{g(x)} g'(x) = \frac{g'(x)}{g(x)}$$

Example

The derivative of

$$f(x) = \ln(1 - x)$$

is (for all $x < 1$)

$$f'(x) = \frac{1}{1 - x}(-1) = \frac{1}{x - 1}$$

- For differentiating the function

$$f(x) = x^x$$

neither the power rule (it requires the exponent to be a constant) nor the rule for exponential functions (it requires the base to be a constant) can be applied.

- Taking natural logarithms of each side gives

$$\ln f(x) = \ln x^x$$

and therefore

$$\ln f(x) = x \ln x$$

- Differentiating both sides with respect to x gives

$$\frac{1}{f(x)} f'(x) = \ln x + x \frac{1}{x}$$

- Noting that $f(x) = x^x$ gives

$$\frac{1}{x^x} f'(x) = \ln x + 1$$

and multiplying both sides by x^x yields

$$f'(x) = x^x (\ln x + 1)$$

5 Single-Variable Optimization

5.1 Introduction

- The points in the domain of f where $f(x)$ reaches a maximum or a minimum are called *extreme points* or *optimal points*.
- Every extreme point (optimal point) is either a *maximum point* or a *minimum point* (exception: $f(x) = a$ with a being a constant).

Definition

If $f(x)$ has the domain D, then

$$\begin{aligned} c &\in D \text{ is a max. point for } f(x) \Leftrightarrow f(x) \leq f(c) \text{ for all } x \in D \\ d &\in D \text{ is a min. point for } f(x) \Leftrightarrow f(x) \geq f(d) \text{ for all } x \in D \end{aligned}$$

- If in the definition a strict inequality applies, then we speak of a *strict maximum point* or a *strict minimum point*.
- If c is a maximum point, then $f(c)$ is called the *maximum value*.
- If d is a minimum point, then $f(d)$ is called the *minimum value*.
- If c is a maximum point of the function f, then it is a minimum point of the function $-f$.
- Therefore, a maximization problem can always be converted into a minimization problem, and vice versa.

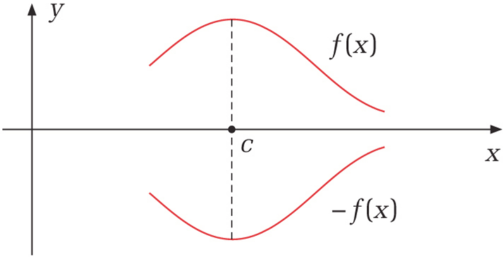


Figure 5-1

- Except for the boundary points of the domain D, every point in D is an *interior point*.
- If f is a differentiable function that has a maximum or minimum at an interior point $c \in D$, then the tangent line to its graph must be horizontal at that point.
- When the tangent line is horizontal, the corresponding point c is called a *stationary point*.

Rule (First-Order Condition)

Suppose that a function f is differentiable in an interval I and that c is an interior point of I. For $x = c$ to be a maximum point for f in I, a necessary condition is that it is a stationary point for f:

$$f'(c) = 0 \qquad \text{(first order condition)}$$

- Figure 5-2 illustrates the meaning of the first-order condition.
- The two stationary points c and d are extreme points.
- However, the first-order condition says nothing about those points of a function that are not differentiable.
- In Figure 5-3 no stationary point exists.
- Points a and b are not interior points.
- The points b and d are extreme points, even though they are not differentiable.

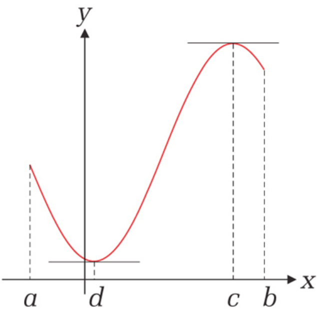


Figure 5-2

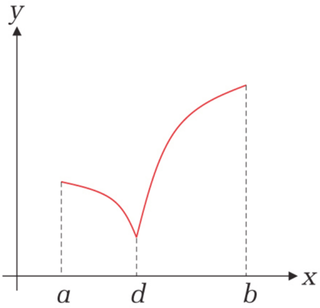


Figure 5-3

- The first-order condition merely states a *necessary* condition for an interior extreme point of a differentiable function.
- Figure 5-4 illustrates that the condition is not *sufficient*.
- It shows three stationary points: x_0, x_1, and x_2.
- Neither of these points is an extreme point.
- At the stationary point x_0 the function f has a *local maximum* (a *local extreme point*).
- At x_1 it has a *local minimum* (another local extreme point).
- x_2 is not a local extreme point.

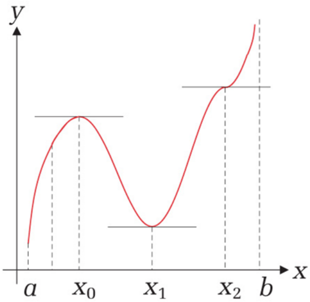


Figure 5-4

5.2 Simple Tests for Extreme Points

- Studying the sign of the derivative of a function f can help to find its maximum or minimum points.

Definition (First-Derivative Test)

If $f'(x) \geq 0$ for $x \leq c$ and $f'(x) \leq 0$ for $x \geq c$, then $x = c$ is a maximum point for f.

If $f'(x) \leq 0$ for $x \leq d$ and $f'(x) \geq 0$ for $x \geq d$, then $x = d$ is a minimum point for f.

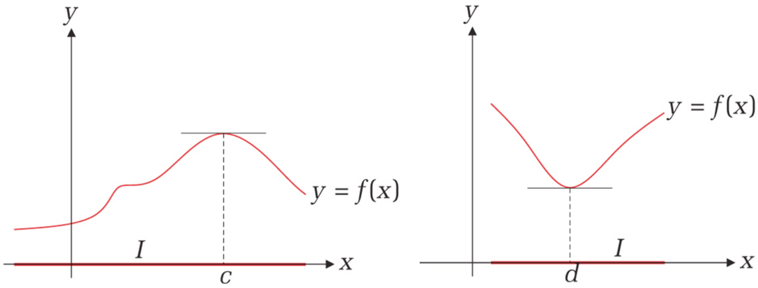


Figure 5-5

Example

The concentration of a drug in the bloodstream t hours after injection is given by the formula

$$c(t) = \frac{t}{t^2 + 4}$$

For finding the time of maximum concentration $c(t)$ must be differentiated with respect to t:

$$c'(t) = \frac{1 \cdot \left(t^2 + 4\right) - t \cdot 2t}{\left(t^2 + 4\right)^2} = \frac{4 - t^2}{\left(t^2 + 4\right)^2} = \frac{(2 - t)\,(2 + t)}{\left(t^2 + 4\right)^2}$$

For $t \geq 0$, the term $(2 - t)$ alone determines the algebraic sign of the fraction. If $t \leq 2$, then $c'(t) \geq 0$, whereas if $t \geq 2$, then $c'(t) \leq 0$. Therefore $t = 2$ is a maximum.

- Recall that

$$\begin{aligned} f''(x) &\geq 0 \text{ for all } x \text{ in } I & \iff && f \text{ is } \textit{convex} \text{ on } I \\ f''(x) &\leq 0 \text{ for all } x \text{ in } I & \iff && f \text{ is } \textit{concave} \text{ on } I \end{aligned}$$

- The first-derivative test is also useful for concave and convex functions.

Rule

Suppose f is a concave (convex) function in an interval I. If c is a stationary point for f in the interior of I, then c is a maximum (minimum) point for f in I.

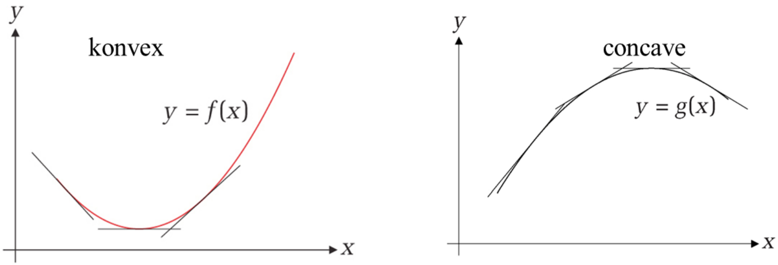


Figure 5-6

5.3 The Extreme Value Theorem

- Recall that stationary points are not necesserily extreme points (Figure 5-4) and that extreme points are not necessarily stationary points (Figure 5-3).
- The following theorem gives a sufficient condition for the existence of a minimum and a maximum.

Rule (Extreme Value Theorem)

Suppose that f is a continuous function over a closed and bounded interval $[a, b]$. Then there exists a point d in $[a, b]$ where f has a minimum, and a point c in $[a, b]$ where f has a maximum, so that

$$f(d) \leq f(x) \leq f(c) \qquad \text{for all } x \text{ in } [a, b]$$

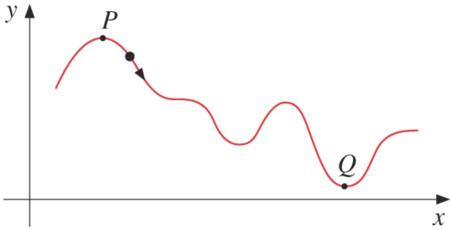


Figure 5-7

- Every extreme point must belong to one of the following three different sets:
  (a) interior points in I where $f'(x) = 0$ (stationary points)
  (b) end points of I (if included in I)
  (c) interior points in I where f' does not exist.
- Points satisfying any one of these three conditions will be called *candidate extreme points*.

- In economics we usually work with functions that are differentiable everywhere. This rules out extreme points of type (c).

Rule

Therefore, the following procedure can be applied to find the extreme points:

1. Find all stationary points of f in (a, b).
2. Evaluate f at the end points a and b and also at all stationary points.
3. The largest function value found in step 2 is the maximum value, and the smallest function value is the minimum value of f in $[a, b]$.

5.4 Local Extreme Points

- So far the chapter discussed *global* optimization problems, that is, all points in the domain were considered without exception.
- In Figure 5-8 c_1, c_2, and b are local maximum points and a, d_1, and d_2 are local minimum points.
- Point d_1 is the global minimum, point b the global maximum.
- The approach to the analysis of global extreme points can be largely adapted to local extreme points. Instead of the domain D only the neighbourhood of a local extreme point must be considered.

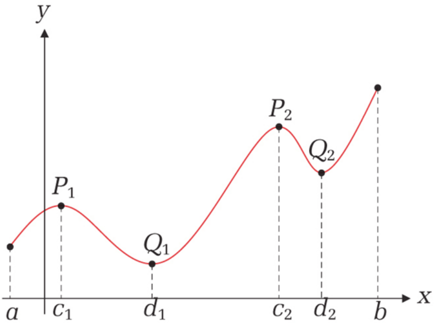


Figure 5-8

5.5 Inflection Points

- Points at which a function changes from being convex to being concave, or vice versa, are called *inflection points*.

Definition

The point c is called an inflection point for the function f if there exists an interval (a, b) about c such that:

(a) $f''(x) \geq 0$ in (a, c) and $f''(x) \leq 0$ in (c, b),
or
(b) $f''(x) \leq 0$ in (a, c) and $f''(x) \geq 0$ in (c, b)

- If c is an inflection point, then we refer to the point $(c, f(c))$ as an inflection point on the graph of f.

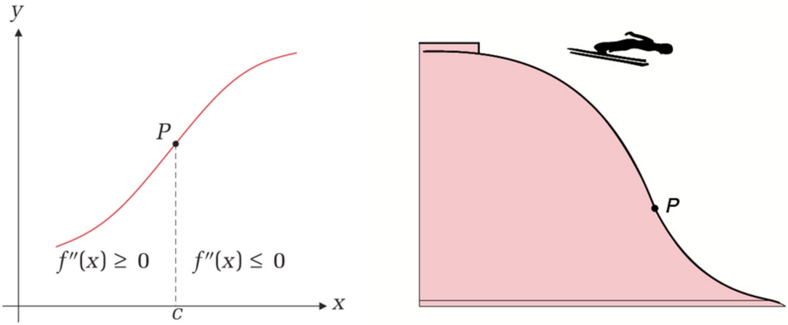


Figure 5-9

Rule (Test for Inflection Point)

Let f be a function with a continuous second derivative in an interval I, and let c be an interior point in I.

(a) If c is an inflection point for f, then $f''(c) = 0$.

(b) If $f''(c) = 0$ and f'' changes sign at c, then c is an inflection point for f.

- Part (a) says that $f''(c) = 0$ is a necessary condition for an inflection point at c.
- However, it is not a sufficient condition. Part (b) says that also a change of the sign of f'' is required.

Example

The function

$$f(x) = x^4$$

has the first derivative

$$f'(x) = 4x^3$$

and the second-order derivative

$$f''(x) = 12x^2$$

Therefore

$$f''(0) = 0$$

but $f''(x)$ does not change sign at $x = 0$.

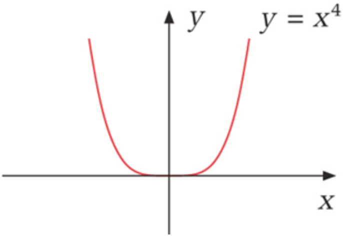


Figure 5-10

Example

The cubic function

$$f(x) = \frac{1}{9}x^3 - \frac{1}{6}x^2 - \frac{2}{3}x + 1$$

has the first derivative

$$f'(x) = \frac{1}{3}x^2 - \frac{1}{3}x - \frac{2}{3}$$

and the second-order derivative

$$f''(x) = \frac{2}{3}x - \frac{1}{3} = \frac{2}{3}\left(x - \frac{1}{2}\right)$$

Therefore $f''(1/2) = 0$ and $f''(x) \geq 0$ for $x \geq 1/2$ and $f''(x) \leq 0$ for $x \leq 1/2$. Hence, $x = 1/2$ is an inflection point for f.

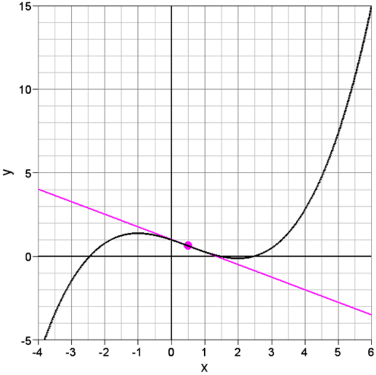


Figure 5-11

6 Functions of Many Variables

6.1 Functions of Two Variables

- For many economic applications functions with more than one independent (or exogenous) variable are necessary.
- With two independent variables x and y the domain D is not a subset of the x-line but a subset of the x-y-plane.

Definition

A function f of two variables x and y with domain D is a rule that assigns a specified number $f(x, y)$ to each point (x, y) in D.

- Often the value of f at (x, y) is denoted by z, so $z = f(x, y)$.
- z is the dependent (or endogenous) variable.
- Unless otherwise stated, the domain of a function defined by a formula is the largest domain in which the formula gives a meaningful and unique value.

Example

The Cobb-Douglas function (with two independent variables) is defined as

$$f(x, y) = Ax^a y^b$$

with A, a, and b being constants. It is often used to describe a production process in which the inputs x and y are transformed into output $z = f(x, y)$. What happens to the output z when both inputs x and y are doubled? A doubling of x and y leads to

$$\begin{aligned} f(2x, 2y) &= A(2x)^a(2y)^b = A2^a2^bx^ay^b \\ &= 2^{a+b}Ax^ay^b = 2^{a+b}f(x, y) \end{aligned}$$

If $a + b = 1$, then a doubling of both inputs x and y leads to a doubling of output z.

Example (continued)

More generally, the Cobb-Douglas function yields

$$\begin{aligned} f(tx, ty) &= A(tx)^a (ty)^b = At^a t^b x^a y^b \\ &= t^{a+b} A x^a y^b = t^{a+b} f(x, y) \end{aligned}$$

For example, if $a + b = 0.7$, then the equation implies that a 10%-increase in inputs ($t = 1.1$) increases output by

$$1.1^{0.7} f(x, y) - 1^{0.7} f(x, y) = \left(1.1^{0.7} - 1\right) f(x, y) = 0.068993\, f(x, y)$$

This is a 6.8993% increase in output.

Definition (Homogeneous Functions)

A function $f(x, y)$ with the property

$$f(tx, ty) = t^q f(x, y) \tag{49}$$

is called a homogeneous function of degree q.

6.2 Partial Derivatives with Two Variables

- For a function $y = f(x)$ the derivative was denoted by

$$\frac{\mathrm{d}y}{\mathrm{d}x} \qquad \text{or} \qquad f'(x)$$

  measuring the function's rate of change as x changes, that is, the number of units that y changes as x changes by one unit.
- For a function $z = f(x, y)$ one may also want to know the function's rate of change as one of the independent variables changes *and the other independent variable is kept constant*.

Example

Consider again the Cobb-Douglas function

$$f(x, y) = Ax^a y^b$$

Changing input x (by Δx) and keeping input y constant changes output by

$$\begin{aligned} f(x + \Delta x, y) - f(x, y) &= A(x + \Delta x)^a y^b - Ax^a y^b \\ &= Ay^b\left((x + \Delta x)^a - x^a\right) \end{aligned}$$

This says that output increases by $Ay^b\left((x + \Delta x)^a - x^a\right)$ units when x is increased by Δx units while y is kept constant.

Definition

If $z = f(x, y)$, then

(i) $\dfrac{\partial z}{\partial x}$ denotes the derivative of $f(x, y)$ with respect to x when y is held constant;

(ii) $\dfrac{\partial z}{\partial y}$ denotes the derivative of $f(x, y)$ with respect to y when x is held constant.

- The derivatives

$$\frac{\partial z}{\partial x} \quad \text{and} \quad \frac{\partial z}{\partial y}$$

are denoted as the *partial derivatives* of the function $z = f(x, y)$.

Definition

The partial derivatives of the function $z = f(x, y)$ at point (x_0, y_0) are given by the formulas

$$\frac{\partial z}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

$$\frac{\partial z}{\partial y} = \lim_{\Delta y \to 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

- To find $\partial z/\partial x$, we can think of y as a constant and can differentiate $f(x, y)$ with respect to x as if f were a function only of x.
- Therefore, the ordinary rules of differentiation can be applied.

Example

The partial derivatives of

$$z = x^3 y + x^2 y^2 + x + y^2 \tag{50}$$

are

$$\begin{aligned} \frac{\partial z}{\partial x} &= 3x^2 y + 2xy^2 + 1 \\ \frac{\partial z}{\partial y} &= x^3 + 2x^2 y + 2y \end{aligned}$$

Example

The partial derivatives of

$$z = \frac{xy}{x^2 + y^2}$$

are (applying the quotient rule)

$$\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{y\left(x^2 + y^2\right) - xy2x}{\left(x^2 + y^2\right)^2} = \frac{y\left(y^2 - x^2\right)}{\left(x^2 + y^2\right)^2} \\
\frac{\partial z}{\partial y} &= \frac{x\left(x^2 - y^2\right)}{\left(x^2 + y^2\right)^2}
\end{aligned}$$

- Some of the most common alternative forms of notation for partial derivatives are

$$
\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{\partial f}{\partial x} = \frac{\partial f(x,y)}{\partial x} = z'_x = f'_x(x,y) = f'_1(x,y) \\
\frac{\partial z}{\partial y} &= \frac{\partial f}{\partial y} = \frac{\partial f(x,y)}{\partial y} = z'_y = f'_y(x,y) = f'_2(x,y)
\end{aligned}
$$

- The variants with $f(x,y)$ are better suited when we want to emphasize the point (x,y) at which the partial derivative is evaluated.

- If $z = f(x, y)$, then $\partial z/\partial x$ and $\partial z/\partial y$ are called *first-order partial derivatives*.

Definition

Differentiating $\partial z/\partial x$ with respect to x and y generates the *second-order partial derivatives*

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x^2} \quad \text{and} \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x \partial y}$$

In the same way, differentiating $\partial z/\partial y$ with respect to x and y generates the *second-order partial derivatives*

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y \partial x} \quad \text{and} \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y^2}$$

Example

The first-order partial derivatives of the function (50) were

$$\frac{\partial z}{\partial x} = 3x^2y + 2xy^2 + 1 \qquad \text{and} \qquad \frac{\partial z}{\partial y} = x^3 + 2x^2y + 2y$$

The second-order partial derivatives are

$$\begin{aligned} \frac{\partial^2 z}{\partial x^2} &= 6xy + 2y^2 \qquad \text{and} \qquad \frac{\partial^2 z}{\partial x \partial y} = 3x^2 + 4xy \\ \frac{\partial^2 z}{\partial y \partial x} &= 3x^2 + 4xy \qquad \text{and} \qquad \frac{\partial^2 z}{\partial y^2} = 2x^2 + 2 \end{aligned}$$

- For most functions $f(x, y)$ it is true that

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$$

- Some of the most common alternative forms of notation for second-order partial derivatives are

$$\begin{aligned} \frac{\partial^2 z}{\partial x^2} &= \frac{\partial^2 f}{\partial x^2} = f''_{xx}(x, y) = f''_{11}(x, y) \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial x \partial y} = f''_{xy}(x, y) = f''_{12}(x, y) \end{aligned}$$

- Also partial derivatives of higher order can be defined.

6.3 Geometric Representation

- A function $z = f(x, y)$ has a graph which forms a surface in three-dimensional space.
- This space has an x-axis, an y-axis, and a z-axis.
- These axes are mutually orthogonal (a 90-degree angle between each of them) – see Figure 6-1.
- The arrows point in the positive direction.
- Any point in (three-dimensional) space is represented by ordered triples of real numbers (x, y, z).
- Figure 6-1 shows the point $P = (x_0, y_0, z_0)$.
- Figure 6-2 shows the point $P = (-2, 3, -4)$.

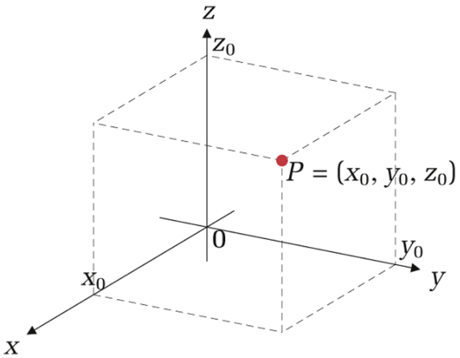


Figure 6-1

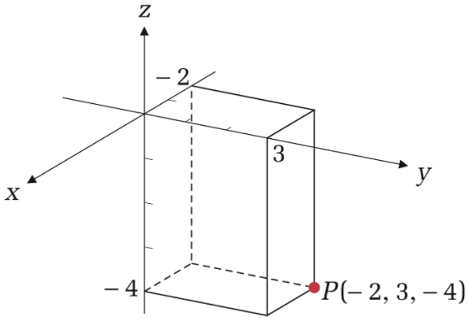


Figure 6-2

- The equation $z = 0$ is satisfied by all points in the coordinate plane spanned by the x-axis and the y-axis. This is called the x-y-plane.
- The x-y-plane is usually thought of as the horizontal plane and the z-axis passes vertically through this plane.
- The x-y-plane divides the space into two half-spaces, one representing all points with $z > 0$ (above the x-y-plane) and the other one representing all points with $z < 0$ (below the x-y-plane).
- The domain of a function $f(x, y)$ can be viewed as a subset of the x-y-plane.

- Suppose $z = f(x, y)$ is defined over a domain D in the x-y-plane.
- The graph of function f is the set of all points $(x, y, f(x, y))$ obtained by letting (x, y) "run through" the whole of D.
- If f is a "nice" function, its graph will be a connected surface in the space, like the graph in Figure 6-3.
- The point $P = (x_0, y_0, f(x_0, y_0))$ on the surface is obtained by letting $f(x_0, y_0)$ be the "height" of f at (x_0, y_0).

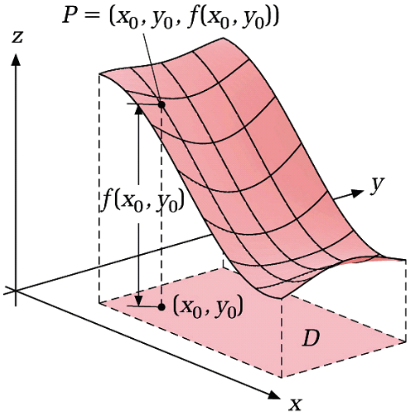


Figure 6-3

- Sometimes a three-dimensional relationship must be represented in two-dimensional space.
- For this purpose, topographical maps use level curves or contours connecting points on the map that represent places with the same elevation level.
- Also for an arbitrary function $z = f(x, y)$ such level curves can be drawn.
- A level curve corresponding to level $z = c$ is obtained by the intersection of the plane $z = c$ and the graph of f.
- In Figure 6-4 the function $z = f(x, y)$ represents a cone (indicated by the red arch) and the plane $z = c$ is indicated by the red framed rectangle.

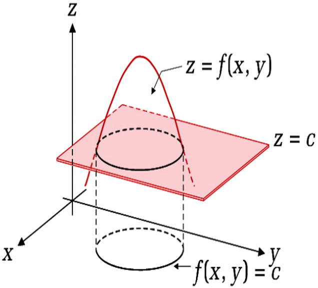


Figure 6-4

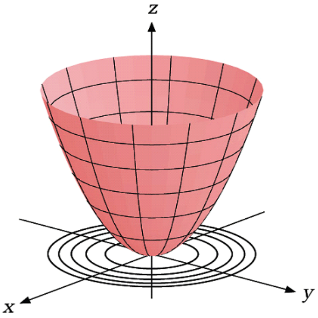
- This level curve consists of points satisfying the equation

$$f(x, y) = c$$

- Finally, the level curve is projected on the x-y-plane.
- This procedure can be done for different levels.
- One obtains a set of level curves projected on the x-y-plane.

Example

Figure 6-5 shows the graph and the level curves corresponding to the function $z = x^2 + y^2$.



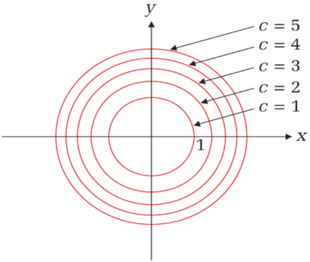


Figure 6-5

Example

Suppose that the output Y of a firm is produced by the inputs capital K and labour L by the following Cobb-Douglas production function:

$$F(K, L) = AK^aL^b$$

with $a + b < 1$ and $A > 0$. Figure 6-6 shows the graph near the origin and the corresponding level curves. In the context of production functions, level curves are called *isoquants*.

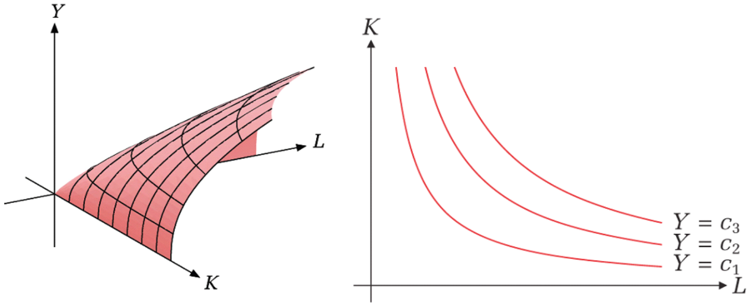


Figure 6-6

- Figure 6-7 depicts the graph of some function $z = f(x, y)$.
- Keeping y_0 fixed, gives the points on the graph that lie on curve K_y.
- Keeping instead x_0 fixed, gives the points on the graph that lie on curve K_x.
- Keeping y_0 and x_0 fixed, gives point P.
- The partial derivative

$$\frac{\partial f(x_0, y_0)}{\partial x}$$

  is the derivative of $z = f(x, y_0)$ with respect to x at the point $x = x_0$, and is therefore the slope of the tangent line l_y to the curve K_y at $x = x_0$.
- This is the "slope of the graph in point P when looking in the direction parallel to the positive x-axis". It is negative.

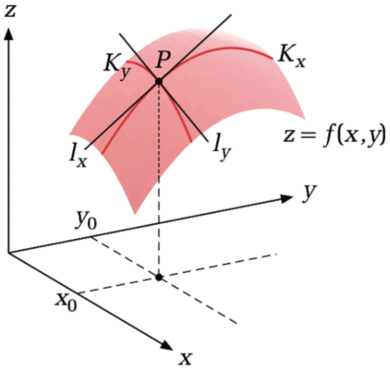


Figure 6-7

- Increasing x above x_0, the partial derivative

$$\frac{\partial f(x, y_0)}{\partial x}$$

decreases (its absolute value increases).

- Therefore, the second-order partial derivative in point $x = x_0$ is negative:

$$\frac{\partial^2 f(x_0, y_0)}{\partial x^2} < 0$$

- The first- and second-order partial derivatives parallel to the y-axis are

$$\frac{\partial f(x_0, y_0)}{\partial y} > 0 \qquad \text{and} \qquad \frac{\partial^2 f(x_0, y_0)}{\partial y^2} < 0$$

6.4 A Simple Chain Rule

- Suppose that

$$z = F(x, y)$$

where x and y both are functions of a variable t, with

$$x = f(t)\ , \qquad\qquad y = g(t)$$

- Substituting for x and y in $z = F(x, y)$ gives the composite function

$$z = F\left(f(t), g(t)\right)$$

- The derivative $\mathrm{d}z/\mathrm{d}t$ measures the rate of change of z with respect to t.

Rule (Chain Rule for One "Basic" Variable)

When $z = F(x, y)$ with $x = f(t)$ and $y = g(t)$, then

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}$$

- This derivative is called the *total derivative* of z with respect to t.
- It is the sum of two contributions:
  1. contribution of x: $\frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}$
  2. contribution of y: $\frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}$

Example

The partial derivatives of

$$z = F(x, y) = x^2 + y^3 \qquad \text{with} \qquad x = t^2 \text{ and } y = 2t$$

are

$$\frac{\partial z}{\partial x} = 2x \qquad \text{and} \qquad \frac{\partial z}{\partial y} = 3y^2$$

Furthermore

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 2t \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = 2$$

So the total derivative is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = 2x \cdot 2t + 3y^2 \cdot 2 = 4tx + 6y^2 = 4t^3 + 24t^2$$

Example (continued)

We can verify the chain rule by substituting $x = t^2$ and $y = 2t$ in the formula for $F(x, y)$ and then differentiating with respect to t:

$$z = x^2 + y^3 = \left(t^2\right)^2 + (2t)^3 = t^4 + 8t^3$$

and therefore

$$\frac{\mathrm{d}z}{\mathrm{d}t} = 4t^3 + 24t^2$$

Example

Consider the Cobb-Douglas agricultural production function

$$Y = F(K, L, T) = AK^a L^b T^c$$

where Y is the size of the harvest, K is capital input, L is labour input, and T is land input. Suppose that K, L, and T are all functions of time t (only one "basic variable"). Then the change in output per unit of time is

$$\begin{aligned} \frac{\mathrm{d}Y}{\mathrm{d}t} &= \frac{\partial Y}{\partial K}\frac{\mathrm{d}K}{\mathrm{d}t} + \frac{\partial Y}{\partial L}\frac{\mathrm{d}L}{\mathrm{d}t} + \frac{\partial Y}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} \\ &= aAK^{a-1}L^b T^c \frac{\mathrm{d}K}{\mathrm{d}t} + bAK^a L^{b-1} T^c \frac{\mathrm{d}L}{\mathrm{d}t} + cAK^a L^b T^{c-1} \frac{\mathrm{d}T}{\mathrm{d}t} \\ &= a\frac{Y}{K}\frac{\mathrm{d}K}{\mathrm{d}t} + b\frac{Y}{L}\frac{\mathrm{d}L}{\mathrm{d}t} + c\frac{Y}{T}\frac{\mathrm{d}T}{\mathrm{d}t} \end{aligned}$$

Example (continued)

Dividing both sides by Y gives

$$\frac{\mathrm{d}Y/\mathrm{d}t}{Y} = a\frac{\mathrm{d}K/\mathrm{d}t}{K} + b\frac{\mathrm{d}L/\mathrm{d}t}{L} + c\frac{\mathrm{d}T/\mathrm{d}t}{T}$$

This is the relative rate of change (percentage change) of output per unit of time.

- Suppose that
$$z = F(x, y)$$
where x and y both are functions of two variables t and s, with
$$x = f(t, s)\ , \qquad\qquad y = g(t, s)$$
- Substituting for x and y in $z = F(x, y)$ gives the composite function
$$z = F\left(f(t, s), g(t, s)\right)$$
- The partial derivative $\partial z/\partial t$ measures the rate of change of z with respect to t, keeping s fixed.
- The partial derivative $\partial z/\partial s$ measures the rate of change of z with respect to s, keeping t fixed.

Rule (Chain Rule for Two "Basic" Variables)

When $z = F(x, y)$ with $x = f(t, s)$ and $y = g(t, s)$, then

$$\begin{aligned} \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \end{aligned}$$

Example

The partial derivatives of

$$z = F(x, y) = x^2 + 2y^2 \qquad \text{with} \qquad x = t - s^2 \text{ and } y = ts$$

are

$$\frac{\partial z}{\partial x} = 2x \qquad \text{and} \qquad \frac{\partial z}{\partial y} = 4y$$

Furthermore

$$\frac{\partial x}{\partial t} = 1, \qquad \frac{\partial x}{\partial s} = -2s, \qquad \frac{\partial y}{\partial t} = s, \qquad \frac{\partial y}{\partial s} = t$$

Example (continued)

Therefore

$$\begin{aligned}
\frac{\partial z}{\partial t} &= 2x \cdot 1 + 4y \cdot s = 2\left(t - s^2\right) + 4ts^2 \\
&= 2t - 2s^2 + 4ts^2 \\
\frac{\partial z}{\partial s} &= 2x \cdot (-2s) + 4y \cdot t = -4\left(t - s^2\right)s + 4t^2 s \\
&= -4ts + 4s^3 + 4t^2 s
\end{aligned}$$

- Suppose that

$$z = F(x_1, ..., x_n)$$

where $x_1, ..., x_n$ are functions of the variables $t_1, ..., t_m$, with

$$x_1 = f_1(t_1, ..., t_m), \quad ... \quad , x_n = f_n(t_1, ..., t_m)$$

- Substituting for $x_1, ..., x_n$ in $z = F(x_1, ..., x_n)$ gives the composite function

$$z = F\left(f_1(t_1, ..., t_m), ..., f_n(t_1, ..., t_m)\right)$$

- The partial derivative $\partial z/\partial t_j$ measures the rate of change of z with respect to t_j, keeping all basic variables t_i with $i \neq j$ fixed.

Rule (Chain Rule for Many "Basic" Variables)

When $z = F(x_1, ..., x_n)$ with

$$x_1 = f_1(t_1, ..., t_m), \quad ... \quad , x_n = f_n(t_1, ..., t_m)$$

then

$$\frac{\partial z}{\partial t_j} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_j} + ... + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_j} \qquad j = 1, 2, ..., m$$

7 Multivariable Optimization

7.1 Introduction

- Figure 7-1 shows on the left hand side the difference between an *interior* and a *boundary point* of some set (domain) S.
- A set is called *open* if it consists only of interior points.
- If the set contains all its boundary points, it is called a *closed* set.

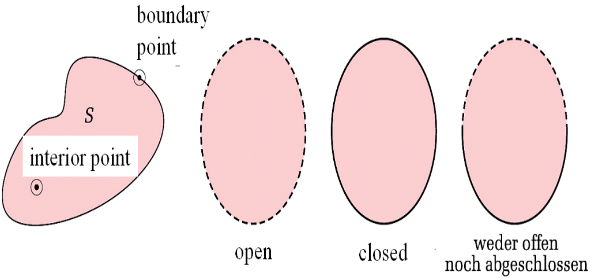

boundary
point
S
interior point
open
closed
weder offen
noch abgeschlossen

Figure 7-1

- The concepts discussed in the context of functions with one independent variable can be applied also in the context of two independent variables.
- Again, we distinguish between
  - local and global extreme points (maxima and minima)
  - interior and boundary (or end) points
  - stationary and non-stationary points.
- We start with local extreme points (Section 7.2). Global extreme points are discussed in Section 7.3.

7.2 Local Extreme Points

Definition (Stationary Points)

Consider the differentiable function $z = f(x, y)$ defined on a set (or domain) S. An interior point (x_0, y_0) of S is a *stationary point*, if the point satisfies the two equations

$$\frac{\partial f(x_0, y_0)}{\partial x} = 0, \qquad \frac{\partial f(x_0, y_0)}{\partial y} = 0 \,. \tag{51}$$

- In Figure 7-1 ("think of it as part of the Himalaya"), there are three stationary points: P, R, and Q.

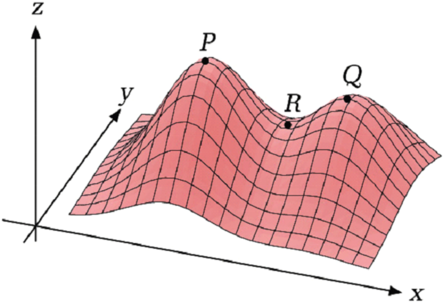


Figure 7-2

Definition

The point (x_0, y_0) is said to be a local maximum point of f in set S if $f(x, y) \leq f(x_0, y_0)$ for all pairs (x, y) in S that lie sufficiently close to (x_0, y_0).

- By "sufficiently close" one should think of a "small" circle with centre (x_0, y_0).
- Points P and Q are *local* maxima.
- Only point P is a *global* maximum.
- Point R is a so-called *saddle point*. This is not an extreme point (more details later).

- Every extreme point of a function $f(x, y)$ must belong to one of the following three different sets:

  (a) an interior point of S that is stationary

  (b) boundary points of S (if included in S)

  (c) interior points in S where $\partial f / \partial x$ or $\partial f / \partial y$ does not exist.

- The following analysis concentrates on variant (a).

Rule (Necessary Condition for a Maximum or Minimum)

A twice differentiable function $z = f(x, y)$ can have a local extreme point (maximum or minimum) at an interior point (x_0, y_0) of S only if this point is a *stationary point*.

- Therefore, the equations (51) are called *first-order conditions* (or FOC's) of a maximum or minimum.
- In Figure 7-3, f attains its largest value (its maximum) at an interior point (x_0, y_0) of S.
- In Figure 7-4, f attains its smallest value (its minimum) at an interior point (x_0, y_0) of S.

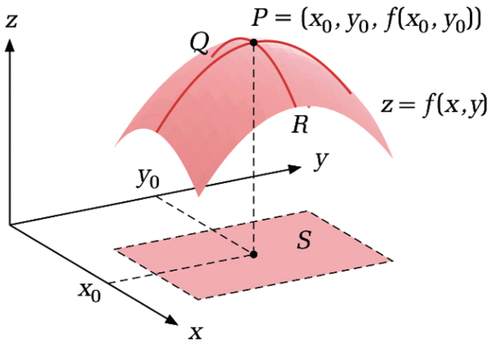


Figure 7-3

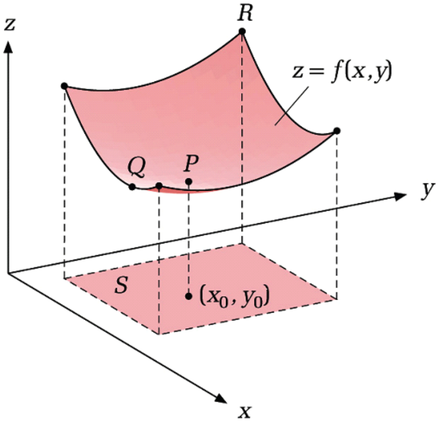


Figure 7-4

Example

The stationary points of the function

$$f(x, y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$$

must satisfy the first-order conditions

$$\begin{aligned}\frac{\partial f}{\partial x} &= -4x - 2y + 36 = 0 \\ \frac{\partial f}{\partial y} &= -2x - 4y + 42 = 0\end{aligned}$$

Multiplying the first condition by $-1/2$ and adding it to the second condition yields:

Example (continued)

$$\begin{aligned} y - 18 - 4y + 42 &= 0 \\ 24 &= 3y \\ y &= 8 \end{aligned}$$

Inserting this result in in the first condition gives

$$\begin{aligned} -4x - 2 \cdot 8 + 36 &= 0 \\ 20 &= 4x \\ x &= 5 \end{aligned}$$

This is the only pair of numbers which satisfies both equations. Therefore, $(x, y) = (5, 8)$ is the only candidate for a local (and global) maximum or minimum.

- Every local extreme point in the interior of set S must be stationary.
- However, not every stationary point in the interior of S is an extreme point.
- The saddle point R of Figure 7-2 was an example.

Definition

A saddle point (x_0, y_0) is a stationary point with the property that there exist points (x, y) arbitrarily close to (x_0, y_0) with $f(x, y) < f(x_0, y_0)$, and there also exist such points with $f(x, y) > f(x_0, y_0)$.

- Figure 7.5 shows another example. This is the graph of the function $f(x, y) = x^2 - y^2$.

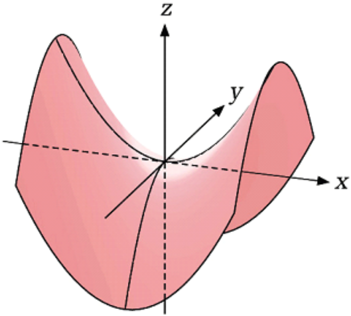


Figure 7-5

Example

The first-order derivatives of the function $f(x, y) = x^2 - y^2$ are

$$\frac{\partial f}{\partial x} = 2x \qquad \text{and} \qquad \frac{\partial f}{\partial y} = -2y$$

Therefore $(0, 0)$ is a stationary point. Moreover, $f(0, 0) = 0$ and for points in the neighbourhood of $(0, 0)$ the function $f(x, 0)$ takes positive values and the function $f(0, y)$ takes negative values. Therefore, $(0, 0)$ is a saddle point.

- Stationary points of a function are either
  - local maximum points,
  - local minimum points,
  - or saddle points.

- For deciding whether a stationary point is a maximum, minimum, or saddle point, we must study the two direct second-order partial derivatives

$$\frac{\partial^2 f}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2} \tag{52}$$

and the two cross second-order partial derivatives

$$\frac{\partial^2 f}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x} \tag{53}$$

Rule (Test for Local Extrema)

Suppose $f(x, y)$ is a twice differentiable function in a domain S, and let (x_0, y_0) be an interior stationary point of S.

(a) If

$$\frac{\partial^2 f}{\partial y^2}\frac{\partial^2 f}{\partial x^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 < 0$$

then (x_0, y_0) is a saddle point.

(b) If

$$\frac{\partial^2 f}{\partial y^2}\frac{\partial^2 f}{\partial x^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 = 0$$

then (x_0, y_0) could be a local maximum, a local minimum, or a saddle point.

Rule (continued)

(c) If

$$\frac{\partial^2 f}{\partial y^2}\frac{\partial^2 f}{\partial x^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 > 0 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x^2} < 0 \tag{54}$$

then (x_0, y_0) is a (strict) local maximum point [Note that (54) automatically implies that $\partial^2 f/\partial y^2 < 0$].

(d) If

$$\frac{\partial^2 f}{\partial y^2}\frac{\partial^2 f}{\partial x^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 > 0 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x^2} > 0$$

then (x_0, y_0) is a (strict) local minimum point.

Example

The first-order conditions of the former example

$$f(x, y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$$

were

$$\frac{\partial f}{\partial x} = -4x - 2y + 36 = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = -2x - 4y + 42 = 0$$

leading to the stationary point $(x, y) = (5, 8)$. The second-order derivatives of all points (x, y) are

$$\frac{\partial^2 f}{\partial x^2} = -4, \quad \frac{\partial^2 f}{\partial y^2} - 4, \quad \frac{\partial^2 f}{\partial x \partial y} = -2 \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x} - 2$$

Example (continued)

Since

$$\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 = 16 - 4 = 12 \geq 0$$

and

$$\frac{\partial^2 f}{\partial x^2} = -4 < 0, \quad \frac{\partial^2 f}{\partial y^2} = -4 < 0$$

the stationary point $(x, y) = (5, 8)$ is a maximum.

7.3 Global Extreme Points

- At most one of the local extreme points is a global maximum and at most one of the local extreme points is a global minimum.

Definition (Convex Set)

A set S in the x-y-plane is *convex* if, for each pair of points P and Q in S, all the line segment between P and Q lies in S.

- The set S in Figure 7-3 is convex.
- For deciding whether a differentiable function $f(x)$ was concave or convex we studied the second-order derivatives.

- For deciding whether a differentiable function $z = f(x, y)$ is concave or convex we must study the two direct second-order partial derivatives

$$\frac{\partial^2 f}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2}$$

and the two cross second-order partial derivatives

$$\frac{\partial^2 f}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x}$$

Definition (Concave or Convex Function)

A twice differentiable function $z = f(x, y)$ is denoted as *concave*, if it satisfies throughout a convex set S the conditions

$$\frac{\partial^2 f}{\partial x^2} \leq 0, \quad \frac{\partial^2 f}{\partial y^2} \leq 0, \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \geq 0,$$

and it is denoted as *convex*, if it satisfies throughout a convex set S the conditions

$$\frac{\partial^2 f}{\partial x^2} \geq 0, \quad \frac{\partial^2 f}{\partial y^2} \geq 0, \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \geq 0.$$

- Figure 7-3 shows a function $f(x,y)$ that is concave in S and Figure 7-4 a function that is convex.

Rule (Sufficient Conditions for a Maximum or Minimum)

Suppose that (x_0, y_0) is an interior stationary point for function $f(x,y)$ defined in a convex set S.

- The point (x_0, y_0) is a (global) maximum point for $f(x,y)$ in S, if $f(x,y)$ is concave.
- The point (x_0, y_0) is a (global) minimum point for $f(x,y)$ in S, if $f(x,y)$ is convex.

Example

In the previous example,

$$f(x,y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$$

we had

$$\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 = 16 - 4 = 12 \geq 0$$

and

$$\frac{\partial^2 f}{\partial x^2} = -4 < 0, \quad \frac{\partial^2 f}{\partial y^2} = -4 < 0$$

Therefore, the function is concave and the stationary point $(x,y) = (5,8)$ is a maximum.

8 Constrained Optimization

8.1 Introduction

- Consider a consumer who chooses how much of the income m to spend on a good x whose price is p, and how much to leave for expenditure y on other goods.
- The consumer faces the budget constraint

$$px + y = m$$

- Suppose that the preferences are represented by the utility function

$$u(x, y)$$

- In mathematical terms, the consumer's *constrained maximization problem* can be expressed as

$$\max u(x, y) \qquad \text{subject to} \qquad px + y = m$$

- This simple problem can be transformed into an unconstrained maximization problem.
- Replace in $u(x, y)$ the variable y by $m - px$ and then maximize this new function

$$h(x) = u(x, m - px)$$

with respect to x.

Example (Consumer Theory)

Suppose that the utility function is

$$u(x, y) = xy \tag{55}$$

and the budget constraint

$$2x + y = 100 \tag{56}$$

Solving the budget constraint for y gives

$$y = 100 - 2x$$

Inserting in the utility function (55) gives

$$u(x, 100 - 2x) = x(100 - 2x)$$

Example (continued)

Differentiating this condition with respect to x gives the first-order condition

$$(100 - 2x) + x(-2) = 0$$

Solving for x gives

$$x = 25$$

and therefore,

$$y = 100 - 2 \cdot 25 = 50$$

Notice that $u''(x) = -4$ for all x. Therefore, $x = 25$ is a maximum.

- However, this substitution method is sometimes difficult or even impossible.
- In such cases the *Lagrange multiplier method* is widely used in economics.

8.2 The Lagrange Multiplier Method

- Suppose that a function $f(x, y)$ is to be maximized, where x and y are restricted to satisfy

$$g(x, y) = c \tag{57}$$

- This can be written as

$$\max f(x, y) \qquad \text{subject to} \qquad g(x, y) - c = 0 \tag{58}$$

- The problem is illustrated in Figure 8.1 for some concave function $f(x, y)$ and some nonlinear constraint $g(x, y) = c$.

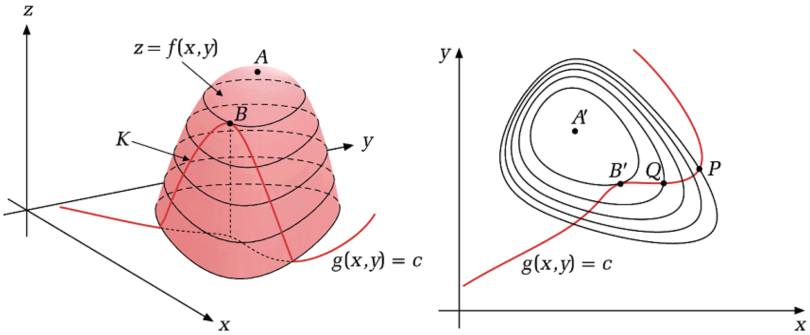


Figure 8-1

- The left hand side diagram shows that the unrestricted maximum is at point A.
- However, the constraint (red and dotted black line in the x-y-plane) implies that only the (x, y)-points on the dotted black line are relevant.
- The restricted maximum value is at point B.
- The right hand side shows the same problem with level curves and the constraint again as a red line.
- Only the x-y-combinations on this red line are available.
- The highest level curve is reached in point B' which corresponds to point B in the left hand diagram.

- The Lagrange multiplier method proceeds in three steps.

Rule

(i) The Lagrange multiplier method introduces a *Lagrange multiplier*, often denoted by λ, and defines the Lagrangian $\mathcal{L}$ by

$$\mathcal{L}(x, y) = f(x, y) - \lambda (g(x, y) - c)$$

The Lagrange multiplier λ should be considered as a constant.

Rule (continued)

(ii) Differentiate $\mathcal{L}$ with respect to x and y, and equate the partial derivatives to 0:

$$\frac{\partial \mathcal{L}(x,y)}{\partial x} = \frac{\partial f(x,y)}{\partial x} - \lambda \frac{\partial g(x,y)}{\partial x} = 0 \quad (59)$$

$$\frac{\partial \mathcal{L}(x,y)}{\partial y} = \frac{\partial f(x,y)}{\partial y} - \lambda \frac{\partial g(x,y)}{\partial y} = 0 \quad (60)$$

(iii) Solve the equations (59) and (60) and the constraint (57) simultaneously for the three unknowns x, y, and λ. These triples (x, y, λ) are the solution candidates, at least one of which solves the problem.

- The conditions (59), (60), and (57) are called the *first-order conditions* for problem (58).

Example (Consumer Theory)

Consider again the utility function (55) and the budget constraint (56). The Lagrangian is

$$\mathcal{L}(x, y) = xy - \lambda(2x + y - 100)$$

The first order conditions are

$$\frac{\partial \mathcal{L}(x, y)}{\partial x} = y - \lambda 2 = 0 \tag{61}$$

$$\frac{\partial \mathcal{L}(x, y)}{\partial y} = x - \lambda = 0 \tag{62}$$

$$2x + y - 100 = 0 \tag{63}$$

Example (continued)

(61) and (62) imply that

$$\begin{aligned} y &= 2\lambda \\ x &= \lambda \end{aligned}$$

Inserting these results in (63) gives

$$2\lambda + 2\lambda = 100$$

and therefore

$$\lambda = 25, \qquad x = 25, \qquad \text{and} \qquad y = 50$$

These are the same results as those derived with the unconstrained maximization.

- Using in the Lagrangian $+\lambda$ instead of $-\lambda$ does not change the results for x and y. Only the sign of λ changes.

Example (Consumer Theory)

Consider again the previous example and use the Lagrangian

$$\mathcal{L}(x,y) = xy + \lambda(2x + y - 100)$$

The first order conditions are

$$\frac{\partial \mathcal{L}(x,y)}{\partial x} = y + \lambda 2 = 0 \tag{64}$$

$$\frac{\partial \mathcal{L}(x,y)}{\partial y} = x + \lambda = 0 \tag{65}$$

$$2x + y - 100 = 0 \tag{66}$$

Example (continued)

(64) and (65) imply that

$$\begin{aligned} y &= -2\lambda \\ x &= -\lambda \end{aligned}$$

Inserting these results in (66) gives

$$-2\lambda + -2\lambda = 100$$

and therefore

$$\lambda = -25, \qquad x = 25, \qquad \text{and} \qquad y = 50$$

These are the same results as those derived with $-\lambda$ in the Lagrangian.

Example (Production Theory)

A firm intends to produce 30 units of output as cheaply as possible. By using K units of capital and L units of labour, it can produce $\sqrt{K}+L$ units of output. Suppose the price of capital is 1 euro and the price of labour is 20 euro. The firm's problem is

$$\min\left(K+20L\right) \qquad \text{subject to} \qquad \sqrt{K}+L=30 \tag{67}$$

The Lagrangian is

$$\mathcal{L}\left(K,L\right)=K+20L-\lambda\left(K^{1/2}+L-30\right)$$

Example (continued)

The first-order conditions are

$$\frac{\partial \mathcal{L}(K,L)}{\partial K} = 1 - \lambda(1/2)K^{-(1/2)} = 0 \tag{68}$$

$$\frac{\partial \mathcal{L}(K,L)}{\partial L} = 20 - \lambda = 0 \tag{69}$$

$$K^{1/2} + L - 30 = 0 \tag{70}$$

(69) gives

$$\lambda = 20 \tag{71}$$

Inserted in (68) yields

$$1 = \frac{20}{2\sqrt{K}}$$

Example (continued)

Therefore,

$$\sqrt{K} = 10 \tag{72}$$

(72) implies that $K = 100$. Inserting (72) in (70) gives

$$L = 20$$

The associated cost is

$$1 \cdot K + 20 \cdot L = 1 \cdot 100 + 20 \cdot 20 = 500$$

Example (Consumer Theory)

A consumer who has a Cobb-Douglas utility function $u(x, y) = Ax^a y^b$ faces the budget constraint $px + qy = m$, where A, a, b, p, and q are positive constants. The consumer's problem is

$$\max Ax^a y^b \qquad \text{subject to} \qquad px + qy = m$$

The Lagrangian is

$$\mathcal{L}(x, y) = Ax^a y^b - \lambda (px + qy - m)$$

Therefore, the first-order conditions are

$$\partial \mathcal{L}(x, y) / \partial x = Aax^{a-1} y^b - \lambda p = 0 \tag{73}$$

$$\partial \mathcal{L}(x, y) / \partial y = Ax^a by^{b-1} - \lambda q = 0 \tag{74}$$

$$px + qy - m = 0 \tag{75}$$

Example (continued)

Solving (73) and (74) for λ yields

$$\lambda = \frac{Ax^a y^{b-1} y}{p} = \frac{Aax^{a-1} y^{b-1} y}{p}$$
$$\lambda = \frac{Ax^{a-1} xy^b}{q} = \frac{Ax^{a-1} xby^{b-1}}{q}$$

Setting the right hand sides equal and cancelling the common factor $Ax^{a-1}y^{b-1}$ gives

$$\frac{ay}{p} = \frac{xb}{q}$$

and therefore

$$qy = px\frac{b}{a}$$

Example (continued)

Inserting this result in (75) yields

$$px + px\frac{b}{a} = m$$

Rearranging gives

$$px = \frac{a}{a+b}m$$

Solving for x yields the following "demand function"

$$x = \frac{a}{a+b}m \cdot \frac{1}{p}$$

Example (continued)

Inserting

$$px = qy\frac{a}{b}$$

in (75) gives

$$\begin{aligned} qy\frac{a}{b} + qy &= m \\ qy &= \frac{b}{a+b}m \end{aligned}$$

and therefore the "demand function"

$$y = \frac{b}{a+b}m \cdot \frac{1}{q}$$

Example (continued)

Suppose that $A = 10$, $a = 0.4$, $b = 0.8$, $p = 2$, $q = 4$, and $m = 1200$. That is, the utility function is $u(x, y) = 10x^{0.4}y^{0.8}$and the budget constraint is $2x + 4y = 1200$. Then our previous results yield the expenditure on x,

$$2x = \frac{a}{a+b}m = \frac{0.4}{1.2}1200 = 400 \ ,$$

and on y,

$$4y = \frac{b}{a+b}m = \frac{0.8}{1.2}1200 = 800 \ .$$

Therefore, the utility maximizing consumption quantities (demands) are $x = 200$ and $y = 200$.

8.3 Interpretation of the Lagrange Multiplier

- Consider the maximization problem

$$\max f(x, y) \qquad \text{subject to} \qquad g(x, y) - c = 0$$

Rule

In a maximization problem with $f'_x > 0$ and $f'_y > 0$, the Lagrange multiplier λ indicates the change in the maximum value of $f(x, y)$ when the constraint $g(x, y) - c = 0$ is relaxed (strengthened) by one unit, that is, when c is increased (decreased) by one unit.

- Consider the minimization problem

$$\min f(x,y) \qquad \text{subject to} \qquad g(x,y)-c=0$$

Rule

In a minimization problem with $f'_x > 0$ and $f'_y > 0$, the Lagrange multiplier λ indicates the change in the minimum value of $f(x,y)$ when the constraint $g(x,y)-c=0$ is strengthened (relaxed) by one unit, that is, when c is increased (decreased) by one unit.

Example (Production Theory)

In a previous example, the problem (67) and the corresponding Lagrangian

$$\mathcal{L}(K, L) = K + 20L - \lambda\left(K^{1/2} + L - 30\right)$$

was considered. The solution was $K = 100$, $L = 20$, and the resulting cost was 500. What is the change in the minimum cost if, instead of 30 units, 31 units are produced (constraint is strengthened)? The new constraint is

$$K^{1/2} + L = 31$$

Again, (69) yields $\lambda = 20$ and (68) yields $K^{1/2} = 10$. Therefore, $K = 100$ and $L = 21$. This implies that the cost increases by one labour unit, that is, by 20 euro. Notice that $\lambda = 20$!

8.4 Several Solution Candidates

- The first-order conditions are necessary conditions for a solution that satisfies the restriction and is in the interior of the domain of (x, y).
- For determining whether the solution is a maximum or a minimum, some ad hoc methods often help.
- These methods are also useful when several solution candidates exist.

Example

The Langrangian associated with the problem

$$\begin{aligned} \max(\min) \qquad & f(x,y) = x^2 + y^2 \\ \text{subject to} \qquad & g(x,y) = x^2 + xy + y^2 = 3 \end{aligned}$$

is

$$\mathcal{L}(x,y) = x^2 + y^2 - \lambda\left(x^2 + xy + y^2 - 3\right)$$

and the first-order conditions are

$$\frac{\partial \mathcal{L}(x,y)}{\partial x} = 2x - \lambda(2x + y) = 0 \tag{76}$$

$$\frac{\partial \mathcal{L}(x,y)}{\partial y} = 2y - \lambda(x + 2y) = 0 \tag{77}$$

$$x^2 + xy + y^2 - 3 = 0 \tag{78}$$

Example (continued)

For $y = -2x$, (76) yields $x = 0$, but (78) yields

$$x^2 + x(-2x) + (2x)^2 - 3 = x^2 - 2x^2 + 4x^2 - 3 = 3x^2 - 3 = 0$$

and therefore, $x = \pm 1$. However, this is a contradiction to $x = 0$.
Therefore $y = -2x$ is not a solution.
Solving (76) for λ yields

$$\lambda = \frac{2x}{2x+y} \qquad \text{(provided } y \neq -2x\text{)}$$

Inserting this value in (77) gives

$$\begin{aligned} 2y - \frac{2x}{2x+y}(x+2y) &= 0 \\ 2y(2x+y) &= 2x(x+2y) \\ y^2 &= x^2 \end{aligned}$$

Example (continued)

Therefore we get

$$y = \pm x$$

Suppose $y = x$. Then (78) yields $x^2 = 1$, so $x = 1$ or $x = -1$. This gives the two solution candidates $(x, y) = (1, 1)$ and $(x, y) = (-1, -1)$, with $\lambda = 2/3$.

Suppose $y = -x$. Then (78) yields $x^2 = 3$, so $x = \sqrt{3}$ or $x = -\sqrt{3}$. This gives the two solution candidates $(x, y) = (\sqrt{3}, -\sqrt{3})$ and $(x, y) = (-\sqrt{3}, \sqrt{3})$, with $\lambda = 2$.

Example (continued)

This leaves the four solutions

$$f(1,1) = f(-1,-1) = 2$$

and

$$f(\sqrt{3}, -\sqrt{3}) = f(-\sqrt{3}, \sqrt{3}) = 6$$

Graphically, $f(x, y)$ is a "bowl standing" on the origin and the constraint $g(x, y) = c$ is an ellipse around the origin. The points furthest away are the maximum points. Here, these are the points $(\sqrt{3}, -\sqrt{3})$ and $(-\sqrt{3}, \sqrt{3})$. The points closest to the origin are the minimum points. Here, these are the points $(1, 1)$ and $(-1, -1)$, see Figure 8.2.

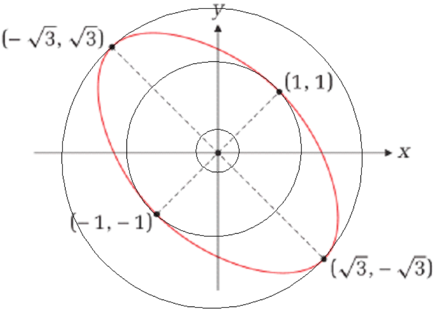


Figure 8-2

8.5 More Than One Constraint

- Suppose that the maximization problem is

$$\max f(x_1, ..., x_n) \qquad \text{subject to} \qquad \begin{cases} g_1(x_1, ..., x_n) = c_1 \\ \vdots \\ g_m(x_1, ..., x_n) = c_m \end{cases}$$

- With each constraint a separate Lagrange multiplier $(\lambda_1, ..., \lambda_m)$ is associated.
- The corresponding Lagrangian is

$$\mathcal{L}(x_1, ..., x_n) = f(x_1, ..., x_n) - \sum_{j=1}^{m} \lambda_j (g_j(x_1, ..., x_n) - c_j)$$

- The solution can be derived from the $n+m$ first-order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}(x_1, \ldots, x_n)}{\partial x_1} &= \frac{\partial f(x_1, \ldots, x_n)}{\partial x_1} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x_1, \ldots, x_n)}{\partial x_1} = 0 \\
&\vdots \\
\frac{\partial \mathcal{L}(x_1, \ldots, x_n)}{\partial x_n} &= \frac{\partial f(x_1, \ldots, x_n)}{\partial x_n} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x_1, \ldots, x_n)}{\partial x_n} = 0 \\
g_1(x_1, \ldots, x_n) &= c_1 \\
&\vdots \\
g_m(x_1, \ldots, x_n) &= c_m
\end{aligned}$$

Example

The Lagrangian of the problem

$$\min \; f(x,y,z) = x^2 + y^2 + z^2 \quad \text{subject to} \quad \begin{cases} x + 2y + z = 30 \\ 2x - y - 3z = 10 \end{cases}$$

is

$$\begin{aligned} \mathcal{L}(x,y,z) &= x^2 + y^2 + z^2 \\ &\quad -\lambda_1 (x + 2y + z - 30) \\ &\quad -\lambda_2 (2x - y - 3z - 10) \end{aligned}$$

Example

The associated first-order conditions are

$$\frac{\partial \mathcal{L}(x, y, z)}{\partial x} = 2x - \lambda_1 - 2\lambda_2 = 0 \tag{79}$$

$$\frac{\partial \mathcal{L}(x, y, z)}{\partial y} = 2y - 2\lambda_1 + \lambda_2 = 0 \tag{80}$$

$$\frac{\partial \mathcal{L}(x, y, z)}{\partial z} = 2z - \lambda_1 + 3\lambda_2 = 0 \tag{81}$$

$$x + 2y + z - 30 = 0 \tag{82}$$

$$2x - y - 3z - 10 = 0 \tag{83}$$

Solving (79) for λ_1 yields

$$\lambda_1 = 2x - 2\lambda_2 \tag{84}$$

Example (continued)

Inserting this value in (80) gives

$$\begin{aligned} 2y - 2\left(2x - 2\lambda_2\right) + \lambda_2 &= 0 \\ 5\lambda_2 &= 4x - 2y \\ \lambda_2 &= \frac{4x - 2y}{5} \end{aligned} \tag{85}$$

Inserting this solution in (84) gives

$$\lambda_1 = 2x - 2\frac{4x - 2y}{5} = \frac{2x + 4y}{5} \tag{86}$$

Example (continued)

Inserting the expressions for λ_1 and λ_2 into (81) gives

$$\begin{aligned} 2z - \frac{2x+4y}{5} + 3\frac{4x-2y}{5} &= 0 \\ 2z + 2x - 2y &= 0 \\ z + x - y &= 0 \end{aligned} \tag{87}$$

(87) gives

$$y = z + x \tag{88}$$

Using this result in (82) yields

$$\begin{aligned} 3y - 30 &= 0 \\ y &= 10 \end{aligned} \tag{89}$$

Example (continued)

Then (88) implies that

$$z = 10 - x \tag{90}$$

Inserting (89) and (90) in (83) gives

$$\begin{aligned} 2x - 10 - 3\left(10 - x\right) - 10 &= 0 \\ -50 + 5x &= 0 \\ x &= 10 \end{aligned} \tag{91}$$

Inserting this result in (90) yields

$$z = 0$$

Example (continued)

Inserting the results for x, y, and z in (85) and (86) gives

$$\begin{aligned} \lambda_2 &= \frac{4 \cdot 10 - 2 \cdot 10}{5} = 4 \\ \lambda_1 &= \frac{2 \cdot 10 + 4 \cdot 10}{5} = 12 \end{aligned}$$

Example (continued)

An easier alternative method to solve this particular problem is to reduce it to a one-variable optimization problem. The constraints are

$$\begin{aligned} x + 2y + z &= 30 &\quad (92)\\ 2x - y - 3z &= 10 &\quad (93) \end{aligned}$$

Multiplying (92) by 2 and then subtracting (93) from the resulting condition yields

$$\begin{aligned} (2x + 4y + 2z) - (2x - y - 3z) &= 60 - 10 \\ 5y + 5z &= 50 \\ y &= 10 - z &\quad (94) \end{aligned}$$

Example (continued)

Inserting this result in (93) gives

$$\begin{aligned} 2x - (10 - z) - 3z &= 10 \\ 2z &= 2x - 20 \\ z &= x - 10 \end{aligned} \tag{95}$$

Inserting (95) in (94) gives

$$y = 10 - (x - 10) = 20 - x \tag{96}$$

Inserting (95) and (96) in $f(x, y, z)$ gives

$$\begin{aligned} h(x) &= x^2 + (20 - x)^2 + (x - 10)^2 \\ &= 3x^2 - 60x + 500 \end{aligned}$$

Example (continued)

The first-order condition is

$$\begin{aligned} h'(x) &= 6x - 60 = 0 \\ x &= 10 \end{aligned}$$

The second-order derivative is

$$h''(x) = 6$$

Therefore, $h(x)$ is convex and $x = 10$ is a minimum. Inserting $x = 10$ in (95) and (96) yields $z = 0$ and $y = 10$. This is the same solution as in the constrained optimization.

9 Matrix Algebra

9.1 Basic Concepts

This chapter covers Section 8.5 (without 8.5.6, 8.5.7, and 8.5.9) of the textbook *Ökonometrie - eine Einführung* (Auer, 2016).

Definition (Matrix)

The *matrix* $\mathbf{A}$

- is a rectangular array of real numbers a_{ij} $(i = 1, 2, ..., Z;\ j = 1, 2, ..., S)$
- that has Z rows and S columns, and therefore, $Z \cdot S$ *elements*

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1S} \\ a_{21} & a_{22} & \cdots & a_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ a_{Z1} & a_{Z2} & \cdots & a_{ZS} \end{bmatrix}$$

- The matrix $\mathbf{A}$ is called a matrix of *order* $(Z \times S)$ or simply a $(Z \times S)$*-matrix*.
- A real number can be interpreted as a (1×1)-matrix.
- Such a matrix is called a *scalar*.
- A matrix with only one row is a *row vector*:

$$\mathbf{a} = \left[\begin{array}{cccc} a_1 & a_2 & \cdots & \end{array} \right]$$

- A matrix with only one column is a *column vector:*

$$\mathbf{b} = \left[\begin{array}{c} b_1 \\ b_2 \\ \vdots \end{array} \right]$$

- A *quadratic matrix* is a matrix with $Z = S$.
- The elements $a_{11}, a_{22} ... a_{ZZ}$ are called the *main diagonal* of a quadratic matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1Z} \\ a_{21} & a_{22} & \cdots & a_{2Z} \\ \vdots & \vdots & \ddots & \vdots \\ a_{Z1} & a_{Z2} & \cdots & a_{ZZ} \end{bmatrix}$$

- If for all elements of a quadratic matrix it is true that $a_{ij} = a_{ji}$, then we speak of a *symmetric matrix*:

$$\mathbf{A} = \begin{bmatrix} a & e & f & g \\ e & b & h & i \\ f & h & c & j \\ g & i & j & d \end{bmatrix}$$

- A *diagonal matrix* is a special case of a symmetric matrix. All its elements except those of the main diagonal are 0:

$$\mathbf{A} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{bmatrix}$$

- A diagonal matrix with $a_{11} = a_{22} = ... = a_{ZZ}$ is a *scalar matrix*:

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

- A scalar matrix with $a_{11} = a_{22} = ... = a_{ZZ} = 1$ is an *identity matrix*:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_4$$

- When all the elements below the main diagonal are 0, then this is an *upper triangular matrix*:

$$\mathbf{A} = \begin{bmatrix} 1 & 7 & 2 \\ 0 & 3 & 9 \\ 0 & 0 & 5 \end{bmatrix}$$

- When all elements above the main diagonal are 0, then this is a *lower triangular matrix*.

- A matrix consisting only of zeros is called a *zero matrix*:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{0}_3$$

- A column vector of zeros is denoted by

$$\mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{o}$$

- A row vector of zeros is denoted by

$$\mathbf{b} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} = \mathbf{o}'$$

Definition (Transposition)

The *transposition* of a matrix is the transformation of a $(S \times Z)$-*matrix* into a $(Z \times S)$-*matrix* by exchanging the rows with the columns.

Example

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \quad \Rightarrow \quad \mathbf{A}' = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

Rule

$$(\mathbf{A}')' = \mathbf{A}$$

- Also vectors can be transposed:

$$\mathbf{a} = \begin{bmatrix} a & b & c \end{bmatrix} \qquad \Rightarrow \qquad \mathbf{a}' = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

9.2 Computing with Matrices

- Two matrices **A** and **B** are identical ($\mathbf{A} = \mathbf{B}$), if they are of the same order and if $a_{ij} = b_{ij}$ $(i = 1, 2, ..., Z; \, j = 1, 2, ..., S)$.

Definition (Summation)

The summation (and subtraction) of two matrices is elementwise and requires that the two matrices are of identical order:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1S} + b_{1S} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2S} + b_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ a_{Z1} + b_{Z1} & a_{Z2} + b_{Z2} & \cdots & a_{ZS} + b_{ZS} \end{bmatrix}$$

Rule

$$\begin{aligned}\mathbf{A}+\mathbf{0} &= \mathbf{A}\\ \mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A}\\ \mathbf{A}'+\mathbf{B}' &= (\mathbf{A}+\mathbf{B})'\end{aligned}$$

- Analogous rules apply to the subtraction of matrices.
- Also three matrices **A**, **B**, and **C** of the same order can be added. Furthermore,

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$$

Example

The following matrices are given:

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} -1 & 2 \\ 4 & 4 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Computing

$$\mathbf{A} - \mathbf{B}' + 2\mathbf{C}$$

gives

$$\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} -1 & 4 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 7 & -1 \\ -1 & 0 \end{bmatrix}$$

Definition (Scalar Multiplication)

In a *scalar multiplication* each element a_{ij} of a matrix $\mathbf{A}$ is multiplied by the scalar λ:

$$\lambda \mathbf{A} = \mathbf{A}\lambda = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1S} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{Z1} & \lambda a_{Z2} & \cdots & \lambda a_{ZS} \end{bmatrix}$$

Example

The following matrix is given:

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$$

A scalar multiplication by $\lambda = 7$ yields

$$7\mathbf{A} = \begin{bmatrix} 7 \cdot 4 & 7 \cdot 3 \\ 7 \cdot 1 & 7 \cdot 2 \end{bmatrix} = \begin{bmatrix} 28 & 21 \\ 7 & 14 \end{bmatrix}$$

The scalar multiplication $\mathbf{A}7$ gives the same result.

Definition (Inner Product)

The *inner product* of the row vector $\mathbf{a}'$ and the column vector $\mathbf{b}$ (each with Z elements) is:

$$\mathbf{a}'\mathbf{b} = a_1 b_1 + a_2 b_2 + \ldots + a_Z b_Z = \sum_{i=1}^{Z} a_i b_i$$

- The result of an inner product is always a scalar.
- The mechanics of calculation: Suppose that $Z = 3$. Then

$$\begin{array}{ccc|c} & & & b_1 \\ & \mathbf{a}'\mathbf{b} & & b_2 \\ & & & b_3 \\ \hline a_1 & a_2 & a_3 & a_1 b_1 + a_2 b_2 + a_3 b_3 \end{array}$$

Example

The following vectors are given:

$$\mathbf{c} = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Computing $\mathbf{c}'\mathbf{d}$ gives

$$\begin{array}{ccc|c}
 & & & 1 \\
\mathbf{c}'\mathbf{d} & & & 2 \\
 & & & 2 \\
\hline
4 & -2 & 3 & 4 \cdot 1 + (-2) \cdot 2 + 3 \cdot 2 = 6
\end{array}$$

- The *multiplication of matrices* requires that the number of columns of the first matrix is identical to the number of rows of the second matrix.
- Let

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \end{aligned}$$

and

$$\mathbf{C} = \mathbf{AB}$$

Definition

The element c_{ij} of matrix $\mathbf{C} = \mathbf{AB}$ is the inner product of row i of matrix $\mathbf{A}$ and column j of matrix $\mathbf{B}$:

$$
\begin{array}{c|c}
 & \mathbf{B} \\
\hline
\mathbf{A} & \mathbf{C}
\end{array}
=
\begin{array}{cc|ccc}
 & & b_{11} & b_{12} & b_{13} \\
 & & b_{21} & b_{22} & b_{23} \\
\hline
a_{11} & a_{12} & c_{11} & c_{12} & c_{13} \\
a_{21} & a_{22} & c_{21} & c_{22} & c_{23}
\end{array}
$$

$$
=
\begin{array}{cc|ccc}
 & & b_{11} & b_{12} & b_{13} \\
 & & b_{21} & b_{22} & b_{23} \\
\hline
a_{11} & a_{12} & a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} & a_{11}b_{13}+a_{12}b_{23} \\
a_{21} & a_{22} & a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} & a_{21}b_{13}+a_{22}b_{23}
\end{array}
$$

Example

The following two matrices are given:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 2 \\ 5 & 6 & 7 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{bmatrix}$$

Calculating $\mathbf{C} = \mathbf{AB}$ gives the following (2×2)-matrix:

$$\begin{array}{ccc|cc} \mathbf{C} & & & 4 & 7 \\ & & & 5 & 8 \\ & & & 6 & 9 \\ \hline 1 & 3 & 2 & 31 & 49 \\ 5 & 6 & 7 & 92 & 146 \end{array}$$

Example

Again, the following two vectors (matrices) are given:

$$\mathbf{c} = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

In a previous example $\mathbf{c}'\mathbf{d}$ was computed. Now $\mathbf{cd}'$ is computed:

$\mathbf{cd}'$	1	2	2
4	4	8	8
−2	−2	−4	−4
3	3	6	6

- The sequence of multiplication is important.
- Right-sided multiplication of matrix **A** by matrix **B** yields **AB** (if the matrices are of coherent orders).
- Left-sided multiplication of matrix **A** by matrix **B** yields **BA** (if the matrices are of coherent orders).
- In general,

$$\mathbf{AB} \neq \mathbf{BA}$$

Example

The following two matrices are given:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 5 & 6 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$$

Calculating $\mathbf{C} = \mathbf{AB}$ and $\mathbf{D} = \mathbf{BA}$ gives the following (2×2)-matrices:

$$\begin{array}{cc|cc} \mathbf{C} & & 1 & 0 \\ & & 1 & 2 \\ \hline 1 & 3 & 4 & 6 \\ 5 & 6 & 11 & 12 \end{array} \qquad \begin{array}{cc|cc} \mathbf{D} & & 1 & 3 \\ & & 5 & 6 \\ \hline 1 & 0 & 1 & 3 \\ 1 & 2 & 11 & 15 \end{array}$$

Rule

Consider a $(Z \times S)$-matrix $\mathbf{A}$. Then

$$\begin{aligned} \mathbf{A}\mathbf{I}_S &= \mathbf{A} \\ \mathbf{I}_Z\mathbf{A} &= \mathbf{A} \\ \mathbf{A}\mathbf{0}_S &= \mathbf{0} \\ \mathbf{0}_Z\mathbf{A} &= \mathbf{0} \end{aligned}$$

Example

The following three matrices are given:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 5 & 6 \end{bmatrix} \qquad \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{0}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Calculating $\mathbf{C} = \mathbf{A}\mathbf{I}_2$ and $\mathbf{D} = \mathbf{0}_2\mathbf{A}$ gives the following (2×2)-matrices:

$$\begin{array}{cc|cc} \mathbf{C} & & 1 & 0 \\ & & 0 & 1 \\ \hline 1 & 3 & 1 & 3 \\ 5 & 6 & 5 & 6 \end{array} \qquad \begin{array}{cc|cc} \mathbf{D} & & 1 & 3 \\ & & 5 & 6 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}$$

Rule

If for the matrices **A**, **B**, **C**, and **D** the respective computations are admissable, then

$$\begin{aligned}
(\mathbf{AB})\,\mathbf{C} &= \mathbf{A}\,(\mathbf{BC}) \\
(\mathbf{A}+\mathbf{B})\,\mathbf{C} &= \mathbf{AC}+\mathbf{BC} \\
\mathbf{A}\,(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC} \\
(\mathbf{A}+\mathbf{B})\,(\mathbf{C}+\mathbf{D}) &= \mathbf{AC}+\mathbf{AD}+\mathbf{BC}+\mathbf{BD} \\
(\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \\
(\mathbf{ABC})' &= \mathbf{C}'\mathbf{B}'\mathbf{A}'
\end{aligned}$$

Rule

Let λ denote a scalar. Then,

$$\lambda \mathbf{AB} = \mathbf{A}\lambda\mathbf{B} = \mathbf{AB}\lambda$$

Definition (Idempotent Matrix)

A quadratic matrix $\mathbf{A}$ for which

$$\mathbf{AA} = \mathbf{A}$$

is denoted as *idempotent*

- The identity matrix $\mathbf{I}_Z$ is an example for an idempotent matrix.

Example

The multiplication $\mathbf{I}_2\mathbf{I}_2$ gives the following result:

$$\begin{array}{cc|cc} & & 1 & 0 \\ & & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}$$

9.3 Rank and Inversion

- Let $\lambda_1, \lambda_2, ..., \lambda_S$ denote real numbers.

Definition (Linear Dependence)

The vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_S$ are linearly dependent, when

$$\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + ... + \lambda_S \mathbf{a}_S = \mathbf{o} \ , \quad \text{where at least one } \lambda_i \neq 0$$

Otherwise, the vectors are linearly independent.

Example

The vectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 2 \\ 0 & -2 & -1 \\ 0 & 2 & 1 \end{bmatrix}$$

are linearly dependent, because for $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = -2$ one gets

$$1 \cdot \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ -2 \\ 2 \end{bmatrix} - 2 \cdot \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- The *column rank* of a matrix **A** is the *maximum* number of *linearly independent* columns.
- The *row rank* of a matrix **A** is the *maximum* number of *linearly independent* rows.
- Column rank and row rank are always identical.
- Therefore, one simply speaks of *the rank* of matrix **A:** $\text{rank}(\mathbf{A})$:

Rule

$$\text{rank}(\mathbf{A}) \leq \min(Z, S)$$

- If

$$\text{rank}(\mathbf{A}) = \min(Z, S)$$

then the matrix has *full* rank.

Rule

$$
\begin{aligned}
\operatorname{rank}(\mathbf{A}') &= \operatorname{rank}(\mathbf{A}) \\
\operatorname{rank}(\mathbf{A}'\mathbf{A}) &= \operatorname{rank}(\mathbf{A}\mathbf{A}') = \operatorname{rank}(\mathbf{A}) \\
\operatorname{rank}(\mathbf{I}_Z) &= Z
\end{aligned}
$$

Definition (Regular and Singular)

A quadratic matrix with full rank is denoted as a *regular matrix*. If the quadratic matrix does not have full rank it is a *singular matrix*.

Definition (Inverse)

To each regular $(Z \times Z)$-matrix $\mathbf{A}$ a matrix $\mathbf{A}^{-1}$ exists that is characterized by the following property:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_Z$$

The matrix $\mathbf{A}^{-1}$ is called the *inverse* of $\mathbf{A}$.

Rules

- If matrix $\mathbf{A}$ is not regular, it does not have an inverse.
- The inverse of a regular matrix $\mathbf{A}$ is also regular.
- Furthermore,

$$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$$

Example

The following two matrices are given:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ -0.5 & 0.5 \end{bmatrix}$$

Calculating $\mathbf{C} = \mathbf{AB}$ gives the following (2×2)-matrix:

$$\begin{array}{cc|cc} \mathbf{C} & & 1 & 0 \\ & & -0.5 & 0.5 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}$$

Therefore, $\mathbf{C} = \mathbf{I}_2$. This implies that $\mathbf{B}$ is the inverse of $\mathbf{A}$: $\mathbf{B} = \mathbf{A}^{-1}$.

Example (continued)

Note that reversing the order of multiplication, $\mathbf{D} = \mathbf{BA}$, gives again

$$\begin{array}{cc|cc} \mathbf{D} & & 1 & 0 \\ & & 1 & 2 \\ \hline 1 & 0 & 1 & 0 \\ -0.5 & -0.5 & 0 & 1 \end{array}$$

Therefore, $\mathbf{A}$ is the inverse of $\mathbf{B}$: $\mathbf{A} = \mathbf{B}^{-1}$. This is a general result. If $\mathbf{B} = \mathbf{A}^{-1}$, then also $\mathbf{A} = \mathbf{B}^{-1}$, and vice versa.

Rules

Computational rules for inverse matrices:

$$\begin{aligned}\left(\mathbf{A}^{-1}\right)' &= \left(\mathbf{A}'\right)^{-1} \\ (\lambda\mathbf{A})^{-1} &= \lambda^{-1}\mathbf{A}^{-1}\end{aligned}$$

- As a consequence,

$$\left[\left(\mathbf{A}'\mathbf{A}\right)^{-1}\right]' = \left[\left(\mathbf{A}'\mathbf{A}\right)'\right]^{-1} = \left[\left(\mathbf{A}'\left(\mathbf{A}'\right)'\right)\right]^{-1} = \left(\mathbf{A}'\mathbf{A}\right)^{-1}$$

Rules

Suppose that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are three arbitrary regular $(Z \times Z)$-matrices. In such a case:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \qquad \text{and} \qquad (\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$$

9.4 Definite and Semidefinite Matrices

- Which of the two matrices

$$\mathbf{B} = \begin{bmatrix} 4 & 0 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 3 \\ 0 & 3 \end{bmatrix}$$

has a "larger value"?

- The difference between the two matrices is

$$\mathbf{A} = \mathbf{B} - \mathbf{C} = \begin{bmatrix} 2 & -3 \\ 3 & 1 \end{bmatrix} \tag{97}$$

- Therefore, no definite answer seems possible.

- A general form of weighting of matrix **A** is the quadratic form

$$
\begin{aligned}
\mathbf{b}'\mathbf{A}\mathbf{b} &= [b_1 \;\; b_2] \begin{bmatrix} 2 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\
&= \begin{bmatrix} 2b_1 + 3b_2 & -3b_1 + b_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\
&= (2b_1 + 3b_2)b_1 + (-3b_1 + b_2)b_2 \\
&= 2b_1 b_1 + b_2 b_2 + 3b_2 b_1 - 3b_1 b_2 \qquad (98) \\
&= 2b_1 b_1 + b_2 b_2 \qquad (99)
\end{aligned}
$$

(98) shows that each element a_{ij} of matrix **A** receives a weight. For example element $a_{21}(=3)$ is weighted by $b_2 b_1$.

- In the numerical example (97), the weighted sum (98) simplifies to expression (99).
- This expression is for all arbitrary values of b_1 and b_2 always positive (except for $b_1 = b_2 = 0$).
- In other words, *regardless of the values of b_1 and b_2*, the quadratic form $\mathbf{b}'\mathbf{A}\mathbf{b}$ yields for the numerical example (97), that is, for the weighted sum (98), always a positive number.
- Therefore, matrix $\mathbf{A}$ is considered as "positive" and, in comparing matrices $\mathbf{B}$ and $\mathbf{C}$, matrix $\mathbf{B}$ is considered as "larger" than $\mathbf{C}$.

- For some general quadratic $(S \times S)$-matrix $\mathbf{A}$, the following definition can be given:

Definition

The *quadratic form* of the quadratic $(S \times S)$-matrix $\mathbf{A}$ is

$$\mathbf{b}'\mathbf{A}\mathbf{b} = \sum_{i=1}^{S} \sum_{j=1}^{S} a_{ij} b_i b_j \tag{100}$$

where $\mathbf{b}' = [b_1 \ b_2 \ \dots \ b_S]$.

- Equation (100) is obtained from:

$$
\begin{aligned}
\mathbf{b}'\mathbf{A}\mathbf{b} &= \begin{bmatrix} b_1 & b_2 & \cdots & b_S \end{bmatrix} \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + \ldots + a_{1S}b_S \\ a_{21}b_1 + a_{22}b_2 + \ldots + a_{2S}b_S \\ \vdots \\ a_{S1}b_1 + a_{S2}b_2 + \ldots + a_{SS}b_S \end{bmatrix} \\
&= b_1(a_{11}b_1 + a_{12}b_2 + \ldots + a_{1S}b_S) \\
&\quad + b_2(a_{21}b_1 + a_{22}b_2 + \ldots + a_{2S}b_S) \\
&\quad \vdots \\
&\quad + b_S(a_{S1}b_1 + a_{S2}b_2 + \ldots + a_{SS}b_S) \\
&= \sum_{i=1}^{S} b_i(a_{i1}b_1 + a_{i2}b_2 + \ldots + a_{iS}b_S) \\
&= \sum_{i=1}^{S} b_i \sum_{j=1}^{S} a_{ij}b_j = \sum_{i=1}^{S}\sum_{j=1}^{S} a_{ij}b_ib_j \ .
\end{aligned}
$$

Definition (Definiteness)

If

$$\begin{array}{lcll} \mathbf{b}'\mathbf{A}\mathbf{b} & > & 0, & \text{matrix } \mathbf{A} \text{ is called } \textit{positive definite} \\ \mathbf{b}'\mathbf{A}\mathbf{b} & < & 0, & \text{matrix } \mathbf{A} \text{ is called } \textit{negative definite} \end{array}$$

If

$$\begin{array}{lcll} \mathbf{b}'\mathbf{A}\mathbf{b} & \geq & 0, & \text{matrix } \mathbf{A} \ \textit{positive semidefinite} \\ \mathbf{b}'\mathbf{A}\mathbf{b} & \leq & 0, & \text{matrix } \mathbf{A} \ \textit{negative semidefinite} \end{array}$$

Rules

- Let $\mathbf{A}$ be an arbitrary $(Z \times S)$-matrix with $\text{rank}(\mathbf{A}) = S$:

$$\mathbf{A}'\mathbf{A} \text{ is always positive definite}$$

- Let $\mathbf{A}$ be a positive definite matrix. Then

$$\mathbf{A}^{-1} \text{ is also positive definite}$$

- For every positive definite $(S \times S)$-matrix $\mathbf{C}$:

$$\text{rank}(\mathbf{C}) = S$$

9.5 Differentiation and Gradient

- Let $\mathbf{a}' = [a_1 \ a_2 \ ... \ a_S]$ be a row vector with S elements and let $\mathbf{b} = [b_1 \ b_2 \ ... \ b_S]'$ be a column vector with S elements.
- Their inner product is

$$\mathbf{a}'\mathbf{b} = a_1 b_1 + a_2 b_2 + ... + a_S b_S = \sum_{i=1}^{S} a_i b_i$$

- The inner product's partial derivative with respect to b_1 is

$$\frac{\partial(\mathbf{a}'\mathbf{b})}{\partial b_1} = a_1$$

Correspondingly,

$$\frac{\partial(\mathbf{a}'\mathbf{b})}{\partial b_S} = a_S$$

Definition (Gradient)

The *gradient* collects all partial derivatives in a single column vector:

$$\frac{\partial(\mathbf{a}'\mathbf{b})}{\partial\mathbf{b}} = \begin{bmatrix} \partial(\mathbf{a}'\mathbf{b})/\partial b_1 \\ \partial(\mathbf{a}'\mathbf{b})/\partial b_2 \\ \vdots \\ \partial(\mathbf{a}'\mathbf{b})/\partial b_S \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_S \end{bmatrix} = \mathbf{a}$$

- Since

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$$

one obtains

$$\frac{\partial(\mathbf{b}'\mathbf{a})}{\partial\mathbf{b}} = \mathbf{a}$$

- Consider the row vector $\mathbf{b}' = [b_1 \ b_2 \ ... \ b_S]$ and the *symmetric* $(S \times S)$-matrix $\mathbf{A}$. The partial derivative of the quadratic form $\mathbf{b}'\mathbf{A}\mathbf{b}$ with respect to b_1 is

$$\begin{aligned}\frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial b_1} &= (a_{11} b_1 + a_{12} b_2 + ... + a_{1S} b_S) + b_1 a_{11} + b_2 a_{21} + b_3 a_{31} + ... + b_S a_{S1} \\ &= 2a_{11} b_1 + (a_{21} + a_{12})\, b_2 + (a_{31} + a_{13})\, b_3 + ... + (a_{S1} + a_{1S})\, b_S\end{aligned}$$

- Since $\mathbf{A}$ is symmetric, we have $a_{ij} = a_{ji}$, and therefore

$$\begin{aligned}\frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial b_1} &= 2a_{11} b_1 + 2a_{12} b_2 + 2a_{13} b_3 + ... + 2a_{1S} b_S \\ &= 2 \sum_{i=1}^{S} a_{1i} b_i\end{aligned}$$

- Analogous results one obtains for b_2, b_3 etc., resulting in the gradient

$$
\begin{aligned}
\frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial \mathbf{b}} &= 2\begin{bmatrix} \sum_{i=1}^{S} a_{1i} b_i \\ \sum_{i=1}^{S} a_{2i} b_i \\ \vdots \\ \sum_{i=1}^{S} a_{Si} b_i \end{bmatrix} \\
&= 2\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1S} \\ a_{21} & a_{22} & \cdots & a_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ a_{S1} & a_{S2} & \cdots & a_{SS} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_S \end{bmatrix} \\
&= 2\mathbf{A}\mathbf{b}
\end{aligned}
$$

Example

Consider the quadratic form of the symmetric Matrix $\mathbf{A}$:

$$\begin{aligned}\mathbf{b}'\mathbf{A}\mathbf{b} &= [b_1\ b_2]\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= 2b_1b_1 + b_2b_2 + 3b_2b_1 + 3b_1b_2 \\ &= 2b_1b_1 + b_2b_2 + 6b_1b_2\end{aligned}$$

The first order partial derivatives with respect to b_1 and b_2 are

$$\begin{aligned}\frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial b_1} &= 4b_1 + 6b_2 \\ \frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial b_2} &= 2b_2 + 6b_1 = 6b_1 + 2b_2\end{aligned}$$

Example (continued)

Therefore, the gradient is

$$\begin{aligned}\frac{\partial(\mathbf{b}'\mathbf{A}\mathbf{b})}{\partial \mathbf{b}} &= \begin{bmatrix} 4b_1 + 6b_2 \\ 6b_1 + 2b_2 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= 2 \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = 2\mathbf{A}\mathbf{b}\end{aligned}$$